2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) | 979-8-3503-9737-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACCS57279.2023.10112932

Classification Of Toddler, Child, Adolescent and Adult for Autism Spectrum Disorder Using Machine Learning Algorithm

V. Kavitha
Department of Computational Intelligence,
School of computing
SRM Institute of Science and Technology
Kattankulathur, Chennai, Tamil Nadu, India.
kavija2010@gmail.com, kv9866@srmist.edu.in
0000-0003-0131-0759

R. Siva
Department of Computational Intelligence,
School of computing
SRM Institute of Science and Technology
Kattankulathur, Chennai, Tamil Nadu, India.
sivar@srmist.edu.in
0000-0002-2006-8753

***Abstract*—A neurodevelopmental syndrome called autism spectrum disorder (ASD) is frequently associated by sensory problems such an excessive or insufficient adaptability to noise, smells, or touch. Our daily lives are relying more and more on Machine Learning (ML). A person with an ASD has lifelong difficulties communicating and interacting socially. This condition begins in childhood and progresses throughout maturity. Therefore, this disorder completely affects a person's life. Reducing symptoms of autism spectrum disorders and improving quality of life for patients with autism, an early diagnosis is essential. Managing the subject's physical and mental health will be substantially helped by the early identification process. The manual screening procedure is more convenient and takes less time. The proposed work is based on an ASD detection mechanism that uses a convolutional neural network and a particle swarm optimization algorithm (PSO-CNN). Initial pre-processing removes missing information from the dataset. To increase the effectiveness of the suggested, we analyse the four underlying techniques, including the SVM, NB, LR, and PSO-CNN with four different dataset types, including ASD Screening Data for Toddlers, Adults, Children, and Adolescents are used for the ASD Prediction analysis. Outcomes strongly indicate that PSO-CNN based prediction models perform more accurately and efficiently on all of these datasets of 99.1% after using various ML approaches with handling missing information.**

Keywords—Autism spectrum disorder, Machine learning, Pre-processing, Particle Swarm Optimization (PSO), Convolutional Neural Network, Support Vector Machine, Naïve Bayes, Classification.

I. INTRODUCTION

ML is an area of study that combines artificial intelligence, mathematics, and other scientific disciplines to produce prediction models from data sets that are accurate and unique. ML is a multidisciplinary study area that uses clever methods to find important hidden patterns that are used in prediction to enhance decision-making. One in one hundred sixty children around the world are estimated by the World Health Organization (WHO) to have an autism spectrum disorder (ASD). There is currently no known treatment for ASD, however one potential solution is to slow the disorder's progression. For ASD individuals and their families, early diagnosis can be crucial to addressing the aforementioned concerns and enhancing quality of life. ASD symptoms often develop at around 36 months old and tend to persist firmly throughout adolescence and adulthood [1].

A neurological condition known as the ASD is characterised by a wide range of symptoms. Social media and communication have several drawbacks, like

- Playing with toys in an improper way
- Unable to make adequate eye contact
- Possibly not wanting to cuddle
- Attachment of inappropriate things
- No pain sensation
- Less susceptible to unexpected noise, light
- Uneasy giggling and laughing

In this article, the remaining sections are organized as follows: Section 1 provides an overview of the issue of ASD and the difficulties the subject's experience. Section 2 outlines the literary works produced by many authors in order to improve accuracy, of autism spectrum disease prediction. The datasets used in this research are described in Section 3.In section 4 dataset using Toddler, Children, Adult, and Adolescents using pre-processing method. Section 5 describes the proposed methodology. Last Section concludes the article.

II. LITERATURE REVIEW

ML Models for Detecting the Early Stage of ASD were demonstrated by Akter et al. [7]. (ASD). Performance was assessed using the Feature Transformation (FT) Method with different classifiers on the converted ASD dataset. The ASD dataset was used to investigate the important ASD characteristics. The measures used to determine a good performance rate include specificity, accuracy, AUROC, Kappa, sensitivity, and log loss. This ASD dataset was predicted with a 97.10% accuracy using a classifier model. ML techniques are effective for spotting ASD in its early stages. In this method, calculation is expensive. According

979-8-3503-9737-6/23/$31.00 ©2023 IEEE

to Sinha et al. [8], electroencephalography (EEG) data can be used to identify ASD.ASD was recognised as a brain disease using both the EEG and a magnetic resonance imaging scan. The system pre-recorded the EEG data and uses the discrete wavelet transform to extract the characteristics in the frequency and time domain. The system gained high accuracy for the features of the time domain with a 92.8% accuracy rate, according to experimentation results.

Uur Erkan et al. [9] method was reported to identify ASD in children, adults, and teenagers. The data was randomly divided into training and test sets, and then we utilised the KNN, SVM and RF methods to categorise the ASD data. A random sample of 100 data points were selected to test the classification techniques. The final outcomes were evaluated using average values.

TABLE 1: METHODS FOR DIAGNOSING ASD

Ref. No.	Author/Year	Merits	Limitation	Algorithms used	Data Collection
[2]	Mohanty et.al. (2021)	Using Decision Tree achieved accuracy is 97%	Detecting ASD cases only for Toddler Not detecting ASD cases on adults, children, adolescents	SVM,KNN,DT, RF	Dataset-UCI Repository
[3]	WalleEt al.(2015)	A subset of autism spectrum behavior was assessed in the ADOS	The overall accuracy of the ASD risk assessment is 98.27%	SVM	-
[4]	T. Y. Rashme et.al (2021)	Logistic Regression gives the highest accuracy with 98%	Detecting ASD cases on adults, children, adolescents Combining all ages should be considered	LR,SVM,RF,KNN	Dataset-UCI Repository
[5]	Mashudi et.al. (2021)	Using SVM achieved accuracy is 97%	Detecting ASD cases only for Adult	SVM,KNN, Bagging, Stacking	Dataset-UCI Repository
[6]	Raj et.al. (2020)	For all three datasets, CNN produced the highest accuracy, with a score of 98.3%.	There is a limited number of samples in the dataset. Rather than discovering biomarkers, this project aims to improve the classification accuracy	Several classification algorithms were evaluated, including LR, SVM, NB, ANN, and CNNs.	Dataset-UCI Repository

The Effectiveness of SVM, KNN, and RF as ASD classification techniques is demonstrated. The performance of RF for classifying ASD data is the best of the three classification methods. The merits and limitation of the data collecting modules and algorithms utilised in the analysis of ASD are summarised in Table 1.

III. DATASET DESCRIPTION

These datasets were gathered from the publicly available UCI repository [10] [11] [12]. Figure 1 displays the datasets description. Three groupings of datasets—AQ-10 for children (Four–Eleven years old), AQ-10 for adolescents (Twelve–Seventeen years old), and AQ-10 for adults—have been created (Eighteen and older). Pre-processing is the conversion carried out prior to the input being supplied into an algorithm. For incomplete and inconsistent data, many pre-processing techniques are used, such as outlier identification, missing values, discretization of data, and null values. The toddler autism screening data set was released in July 2018 via the Kaggle data repository. It has been active since November 2018 and currently has six unique authors. The QCHAT-10 toddler ASD screening instrument [6] provides the basis for the data set. It includes 1054 instances in all.

The UCI Machine Learning Respiratory was the source of the datasets, notably the ASD Screening Data for Adult, Children, and Adolescent. There are 21 attributes and 704 cases in the dataset for adults, 292 instances and 21 attributes in the dataset for children and 104 instances in the dataset for adolescents. Table 2 lists the dataset's common features [13] [14].

There are 10 binary attributes representing the screening questions in all datasets (AID11 to AID20).

Table 3 shows Information on how variables map to the Q-Chat-10 –Toddler screening techniques used in this study [13] [14].

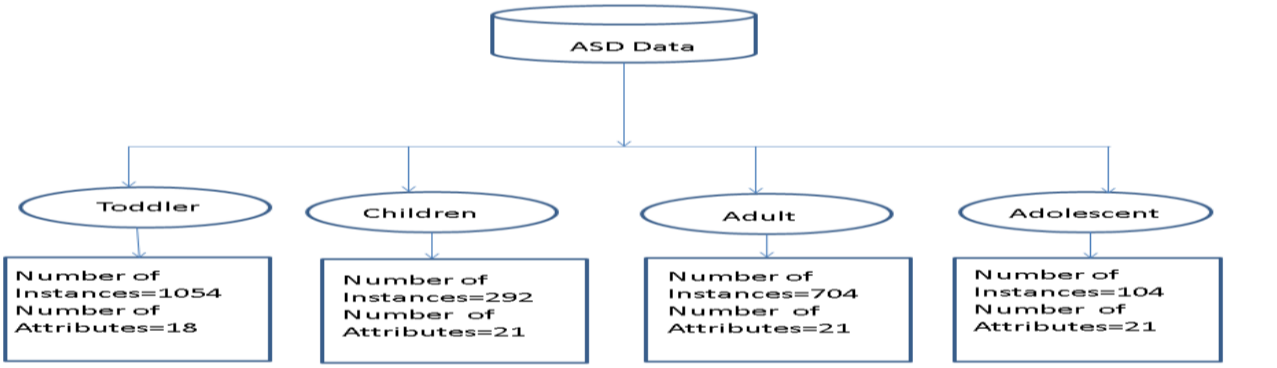

Figure 1. Dataset description of Autism Spectrum Disorder

TABLE 2 : LIST OF ATTRIBUTES

Attr ID	Attributes	Type	Summary
AID1	Age	Number	Age was determined using years
AID2	Gender	String	Male or Female
AID3	Nationality	String	Text-based list of all countries
AID4	Family member with PDD	Boolean (yes or no)	Whether or not any family members had PDD
AID5	Location of residence	String	List of countries where the user lives
AID6	Born with jaundice	Boolean (yes or no)	Whether the patient suffered from jaundice by birth
AID7	Screening application	Boolean (yes or no)	Whether or not a user is using a screening application
AID8	Test completion	String	Who has completed the experiment
AID9	Screening test type	String	Emotion, communication, behaviour and social skills
AID10	Screening Score	Integer	Grade received based on screening technique scoring mechanism
AID11-20	Screening type questions	Binary (0, 1)	Answers based on the screening method
AID21	Class/ASD	Yes/No	Classifying ASD or Non ASD

TABLE 3 : INFORMATION ON HOW VARIABLES MAP TO THE Q-CHAT-10 –TODDLER SCREENING TECHNIQUES

Attr ID	Features of Q-chat-10-Toddler
AID11	When you call your child's name, does he/she look at you?
AID12	Does your child make eye contact easily with you?
AID13	Is your child pointing to express a desire? (For instance, an out-of-reach toy)
AID14	Does your child indicate that they share your interests?
AID15	Does your kid act out? (For instance, tend to dolls or use a toy phone)
AID16	Do they follow where you gaze when you look?
AID17	Does your youngster express an interest in comforting others when they see you or another family member clearly upset? (Examples include hugging them and petting their hair)
AID18	How would you describe the initial words of your child?
AID19	What kind of hand motions does your kid make?
AID20	Does your toddler seem to be staring at nothing but nothing?

IV. DATA PRE-PROCESSING

The raw data is transformed into a format that is comprehensible and meaningful through data preparation. Because there are many null values and errors, the real-world data are inconsistent and lacking. Various data pre-processing methods, such as outlier detection, data reduction, data discretization, and handling missing values,

are used to handle inconsistent and incomplete data. Detailed data pre-processing is done in this step before deep analysis and classification. There are not many records with null values in the ASD datasets. Missing data can be handled by simply deleting the records. The Random Forest technique has been used to address the issues with missing values in these datasets.

Moreover, there are other attributes that are unrelated to ASD and indicate meta-information (e.g., who performed the test, used the app previously). Nonetheless, the datasets must be cleaned up or otherwise prepared before classification [4] [15].

The data reduction is used to reduce the amount of storage space required for the data. Outlier detection uses statistical methods to predict the distribution of datasets. A data discretization method is used to minimize the attribute values

The purpose of this study is to make an early prediction of ASD problems in toddlers, kids, adolescents, and adults. Figure 2 depicts a model that has been proposed for this task, which involves data Pre-processing by using feature selection. The datasets contain some missing data. An "ethnicity," "who is completing the test" and "age" variables are Subjects of missing data. Table 4 displays the missing values in the dataset.

Using Table 4, Figure 3 shows missing values in the dataset.

V. Proposed Methodology

The suggested workflow, which entails pre-processing of the data, training and testing using particular models, assessment of the results, and ASD prediction, is depicted in Figure 4.

A. SVM

One of the most well-liked methods for supervised learning is the Support Vector Machine. The SVM algorithm seeks to produce the optimal decision boundary or line.

B. NB

It is a supervised learning algorithm. It is mostly employed in text categorization with a large training set.

C. LR

Logistic regression is a type of supervised learning. The probability that an event will occur in a binary (yes/no) fashion is determined using this method.

D. CNN

One of the deep learning methods is CNN. A type of synthetic neural network is called a convolutional neural network. one input layer, one output layer, and numerous more layers are present in a CNN model.

E. PSO

PSO is a meta-heuristic technique developed by Dr. Eberhart and Dr. Kennedy in 1995 [16]. By comparing SVM, LR, NB, and CNN, the PSO algorithm was used to complete the feature selection phase, evaluating the best set of features and assisting in achieving the highest accuracy.

TABLE 4 : MISSING VALUES IN THE DATASET

S.No	Name	All	Missing value
	ASD Screening Data for Toddler Data Set	1054	1054
2.	ASD Screening Data for Children Data Set	292	249
3.	ASD Screening Data for Adolescence Data Set	104	98
4.	ASD Screening Data for Adult Data Set	704	609

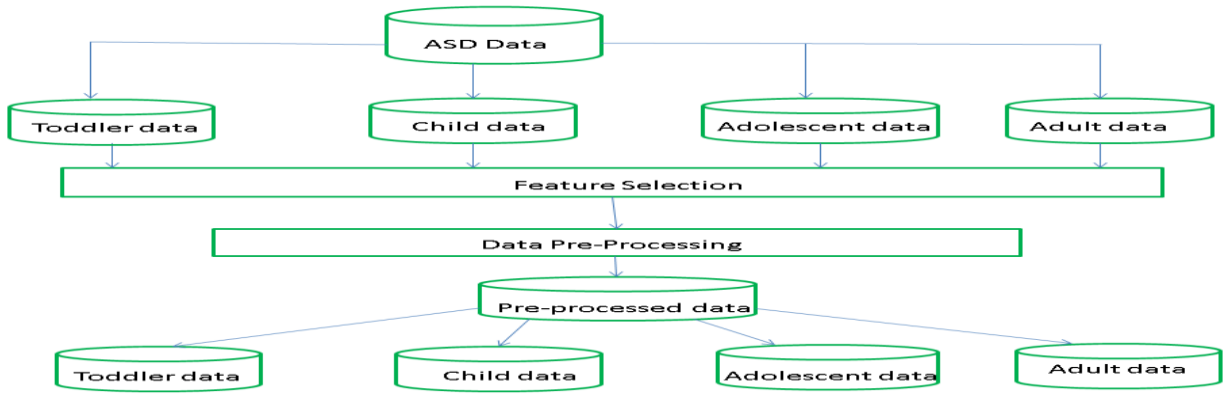

Figure 2. Proposed framework for Pre- processing steps

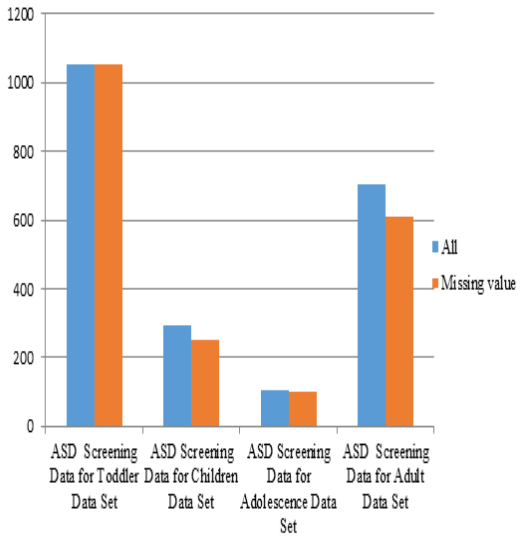

Figure 3. Dataset description of Autism Spectrum Disorder

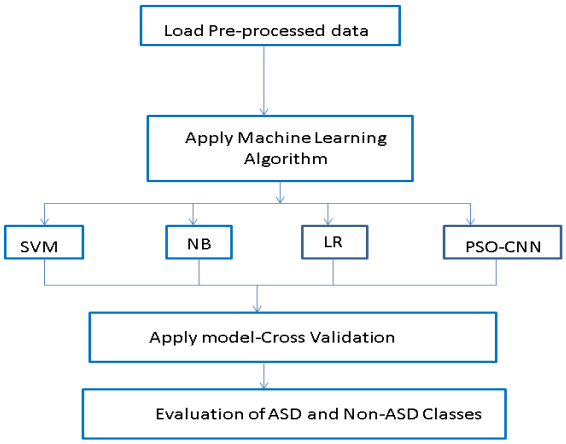

Figure 4. Work flow of proposed method

VI. Results And Discussion

Accuracy, sensitivity, and specificity are measures that are used in the performance evaluation. Methods such as PSO-CNN, SVM, LR, and NB are used for comparative analysis. An "Actual" and "Predicted" dimension, along with TPO, TNE, FPO and FNE on both dimensions, make up a confusion matrix.

$$\text{Accuracy} = \frac{TPO+TNE}{(TPO+TNE+FPO+FNE)}$$

$$\text{Specificity} = \frac{TNE}{TNE+TPO}$$

$$\text{Sensitivity} = \frac{TNE}{TNE+FNE}$$

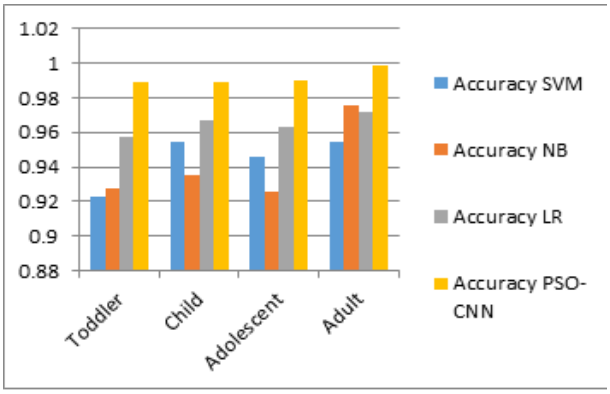

Figure 5. Comparative analysis of accuracy

TABLE 5 : ACCURACY- OVERALL RESULTS FOR ASD

Result	Classifier	Toddler	Child	Adolescent	Adult
Accuracy	SVM	0.9225	0.9545	0.9455	0.9545
	NB	0.9275	0.9356	0.9255	0.9756
	LR	0.9571	0.9668	0.9632	0.9616
	PSO-CNN	0.9894	0.9889	0.9898	0.9987

VII. Conclusion

Four different dataset types, including ASD Screening Data for Adult, Children, Adolescent, and Toddler are used for the ASD prediction model. Before being fed to the algorithm, the collected data from the repository are Pre-processed. The following metrics are used to assess performance: precision, sensitivity, and specificity.

The accuracy rates of various techniques, including LR, SVM, NB, and the proposed PSO-CNN are shown in Figure 5. This comparison analysis shows that the proposed PSO-CNN approach has an accuracy rate of 98.53% on all four datasets. Our suggested model provides performance rates of accuracy in toddlers, children, adolescents, and adults that are, correspondingly, 98.9%, 98.8%, 98.9%, and 99.8%. According to this investigation, the suggested PSO-CNN method outperforms other methods in terms of accuracy. Table 5 shows the accuracy rate of various methods such as LR, SVM, NB, and the proposed PSO-CNN.

In the future, we will assess more information enhancing the ability to identify neurodevelopmental abnormalities like ASD. The lack of enough ASD datasets and restricted access are the key drawbacks of this study. We

also noted that an early ASD diagnosis is definitely achievable. The early diagnosis of ASD will be highly accurate if the ASD datasets data is complete. Furthermore, we would like to analyse brain signals using EEG-fNIRS and EEG-fMRI to develop a more robust autism diagnosis using multi-modality methods./

ABBREVIATIONS

SVM-Support Vector Machine
DT-Decision Tree
RF-Random Forest
PSO-Particle Swarm Optimization
CNN-Convolutional Neural Network
PDD-Pervasive Development Disorders
NB-Naïve Bayes
ANN-Artificial Neural Network
LR-Logistic Regression
TPO-True Positive
TNE-True Negative
FPO-False Positive
FNE-False Negative
UCI-University of California Irvine Machine Learning Repository.

REFERENCES

[1] M. I. Al-Hiyali, N. Yahya, I. Faye, Z. Khan, and K. Alsaih, "Classification of BOLD FMRI Signals using Wavelet Transform and Transfer Learning for Detection of Autism Spectrum Disorder," 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2021, pp. 94-98, doi: 10.1109/IECBES48179.2021.9398803.

[2] Mohanty, Ashima & Patra, Krishna & Parida, Priyadarshan. (2021). Toddler ASD Classification Using Machine Learning Techniques. International Journal of Online and Biomedical Engineering (iJOE). 17. 156. 10.3991/ijoe.v17i07.23497.

[3] J. A. Kosmicki, V. Sochat, M. Duda, and D. P. Wall. (2015) "Searching for a minimal set of behaviours for autism detection through feature selection-based machine learning." Translational psychiatry, 5(2): e514

[4] T. Y. Rashme, L. Islam, A. A. Prova, and S. Jahan, "Autism Screening Disorder: Early Prediction," 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1-6, doi: 10.1109/GUCON50781.2021.9573547.

[5] Mashudi, Nurul, Ahmad, Norulhusna Noor, Norliza 2021/09/01,743,Classification of adult autistic spectrum disorder using machine learning approach, Volume -10,DOI-10.11591/ijai.v10.i3.pp743-751,IAES International Journal of Artificial Intelligence (IJ-AI).

[6] Raj, Suman and Masood, Sarfaraz. (2020). "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques." Procedia Computer Science. 167. 994-1004. 10.1016/j.procs.2020.03.399.

[7] T. Akter et al., "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," in IEEE Access, vol. 7, pp. 166509-166527, 2019, doi: 10.1109/ACCESS.2019.2952609.

[8] Sinha, T., Munot, M.V. and Sreemathy, R., 2022. An efficient approach for detection of autism spectrum disorder using electroencephalography signal. IETE Journal of Research, 68(2), pp.824-832

[9] Erkan, Uğur & Thanh, Dang. (2019). Autism Spectrum Disorder Detection with Machine Learning Methods. 15. 297 - 308. 10.2174/2666082215666191111121115.

[10] Fadi Thabtah, "Autistic Screening Adult Data Set," UCI machine learning repository, 2017. [Online].Available: https://archive.ics.uci.edu/ml.

[11] Fadi Thabtah, "Autistic Spectrum Disorder Screening Data for Adolescent Data Set," UCI machine learning repository, 2017. [Online].Available: https://archive.ics.uci.edu/ml

[12] Fadi Thabtah, "Autistic Spectrum Disorder Screening Data for Children Data Set," UCI machine learning repository, 2017. [Online].Available: https://archive.ics.uci.edu/ml.

[13] Thabtah F. Autism Spectrum Disorder screening: Machine learning adaptation and DSM-5 fulfillment. ICMHI '17 Proceedings of the 1st International Conference on Medical and Health Informatics 2017; Taichung City, Taiwan May 20-22.

[14] Thabtah F. ASDTests. A mobile app for ASD screening. [Internet]. 2017 [cited 2018 Dec 20]. Available from: www.asdtests.com

[15] Thabtah F. ASD Dataset. https://fadifayez.com/autism-datasets/. Accessed 13 Jan 2020.

[16] Sahu, S. K. & Verma, P. (2022). Classification of Autistic Spectrum Disorder Using Deep Neural Network with Particle Swarm Optimization. International Journal of Computer Vision and Image Processing (IJCVIP), 12(1), 1-11. http://doi.org/10.4018/IJCVIP.290398.