Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

A Comprehensive Study of Machine Learning Algorithms to Predict Autism Spectrum Disorder (ASD)

Ms. Pournima P Bhangale
University Department of Information Technology, University of Mumbai, Mumbai, India
Email: ppbhangale@yahoo.com
Dr. Rajendra B Patil
University Department of Information Technology, University of Mumbai, Mumbai, India
Email: patilrajendrab@gmail.com

ABSTRACT

Abstract: Autism Spectrum Disorder (ASD) is a multifaceted neurodevelopmental condition marked by a diverse set of symptoms concerning social communication, limited interests, and repetitive actions. Primary diagnosis and intervention are vital for improving the quality of life for those with ASD and their families. In this study, we conduct a comprehensive investigation into different prediction algorithms for ASD. We collect diverse datasets, perform data preprocessing, employ feature selection and engineering techniques, and evaluate the performance of various algorithms like Random Forest(RF), Naive Bayes(NB), Support Vector Machines(SVM). Our findings shed light on the strengths and limitations of different algorithms in predicting ASD, contributing to improved diagnostic and predictive capabilities in ASD research and clinical practice.

Keywords: **Autism Spectrum Disorder, Machine Learning(ML)Algorithm, Naive Bayes(NB), Random Forest(RF), Support Vector Machines(SVM)**

Date of Submission: 11 October 2023 Date of Acceptance: 15 November 2023

1. Introduction

1.1. Background on autism spectrum disorder

Autism spectrum disorder (ASD) is a developmental disorder characterized by difficulties in social interaction and communication, as well as restricted and repetitive patterns of behavior, interests, or activities. It is called a "spectrum" disorder because its symptoms can vary widely in severity and presentation. ML algorithms have emerged as powerful tools for predicting ASD based on various data sources, including clinical, behavioural, and neuroimaging data. ASD usually begins in early childhood, with noticeable signs and symptoms typically appearing before the age of three. It can affect individuals in different ways, leading to significant differences in their abilities, challenges, and strengths. The exact causes of ASD are still not well understood, but it is believed to be a combination of genetic and environmental factors. Some research suggests that certain genes may raise the risk of developing autism, but multiple genes are likely involved, along with other aspects such as prenatal and early childhood environmental influences[1].

Several of the indications of ASD encompass challenges in social interactions, such as issues with maintaining eye contact, comprehending or responding to social cues, and forming relationships. Communication difficulties can manifest as slow or limited speech, trouble initiating or sustaining conversations, and repetitive or peculiar speech patterns.

There have been several technological contributions to Autism Spectrum Disorder (ASD) that aim to enhance communication, improve social skills, and assist with learning and developmental challenges. Some of these are Augmented and Alternative Communication (AAC) devices, Social Skills Training Apps, Virtual Reality (VR) and Augmented Reality (AR), Assistive Technology for Learning, Wearable Devices and Sensors, Telehealth and Teletherapy. These technological contributions have the potential to significantly improve the quality of life for individuals with ASD, enhance their communication and social skills, support their learning and development, and provide new avenues for intervention and therapy. Some of the limitations of these technological contributions are limited availability and accessibility, lack of personalization, generalizability of skills, and ethical concerns.

1.2. Importance of early detection and intervention

Early detection and intervention are crucial for various reasons. Early detection enables timely intervention and treatment, increasing the chances of successful outcomes. Many health conditions, such as cancer, heart disease, and mental health disorders, are more manageable when detected in their early stages. Early detection can prevent the progression of diseases and conditions. By identifying health issues early, healthcare professionals can initiate appropriate interventions, such as lifestyle changes, medication, or surgical procedures, to halt or slow down the progression. Early detection and intervention can lead to cost savings in the long run. By addressing health problems at an early stage, individuals can avoid more extensive and costly treatments later on. Early intervention also reduces the need for emergency care, hospitalizations, and other expensive medical interventions. Early detection significantly improves survival rates for several conditions. For example, early screening tests for cancer, such as mammography and colonoscopy, can detect abnormalities before they develop into advanced stages, increasing the chances of successful treatment and long-term survival. [2]

Early detection and intervention play a critical role in preventing, managing, and treating a range of health conditions. They not only save lives but also improve quality of life and reduce the burden on healthcare systems.

1.3. Role of machine learning in autism detection

Machine learning can play a significant role in autism detection by providing accurate and efficient ways to analyze and interpret large amounts of data. ML algorithms can analyze data sources such as medical records, genetic information, behavioral patterns, and sensory data to detect early signs of autism. Machine learning can assist healthcare professionals in autism forecasting based on various risk factors and historical data. By considering multiple variables and indicators, ML algorithms can provide objective insights and assist in determining whether an individual has autism. These predictive models can assist in guiding interventions and support early intervention strategies to improve outcomes. Machine learning can analyze individual characteristics, response patterns, and treatment histories to develop personalized treatment plans. By considering data from diverse datasets, including therapy effectiveness, medical history, and genetic information, machine-learning models can identify new patterns or genetic variations of autism detection and optimize treatment recommendations for each individual with autism. Also, it can analyze sensory data collected from wearable devices, such as accelerometers or electrodermal sensors, to identify patterns and deviations.

Overall, machine learning can assist in automating and improving the process of autism detection, providing more accurate and efficient diagnosis, early intervention, and personalized treatment strategies. It has the potential to revolutionize autism research and improve outcomes for individuals with autism.[3..9]. Machine learning (ML) is playing an increasingly crucial role in advancing the understanding of autism spectrum disorder (ASD), facilitating earlier detection, and contributing to the development of more effective treatments and support approaches for autistic individuals. ML offers powerful computational techniques to analyze and interpret vast amounts of data associated with ASD, providing valuable insights into the complex patterns and characteristics of the disorder.

2. Literature Review

2.1. Existing methods for autism detection and their limitations

Current methodologies for autism recognition exhibit diversity and encompass a fusion of behavioral observations, standardized assessments, and the deployment of questionnaires. However, it is imperative to note that there exists no definitive medical diagnostic test or instrument dedicated to autism diagnosis. Instead, clinicians harness these techniques to accrue pertinent information and facilitate well-informed diagnostic judgments[50]. The Diagnostic and Statistical Manual of Mental Disorders (DSM)[10] is a diagnostic framework, used for autism diagnosis. The Modified Checklist for Autism in Toddlers (M-CHAT)[11,45] is a widely adopted screening tool. These assessments can identify potential indicators of autism but they fall short of conferring diagnostic status and are susceptible to both false positives and false negatives[12..20]. Clinical practitioners rely on direct observation of an individual's conduct and developmental trajectory to discern indications of autism[21..25]. Questionnaires and interviews with parents or caregivers provide a valuable resource for getting insights into an individual's behavioral patterns, social interactions, and developmental history.Functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) are also used to explore brain activity and connectivity in individuals with the autism spectrum[26..27,36].

Numerous limitations beset the existing approaches to autism detection. These include the diversity of symptoms, subjectivity, fluctuations in diagnostic criteria, reliance on external observations susceptible to bias, a restricted viewpoint, limited diagnostic specificity, and resource-intensive procedures.

2.2. Previous studies utilizing machine learning algorithms for autism detection

Mamata Lohar et al. [28] assessed the effectiveness of various ML algorithms in the early examination of ASD among diverse age groups. To aid radiologists and neurologists in analyzing sMRI and fMRI images, a computerized support system was developed in which an individual is identified as ASD or a typically developing individual. Employing Principal Component Analysis (PCA) and forward feature selection, they identified the optimal set of features for various classifiers, including K-Nearest Neighbors (K-NN), SVM, NB, Decision Tree(DT), RF, AdaBoost, and Logistic Regression(LR). Notably, their findings revealed that SVM with PCA achieved the highest classification accuracy for one dataset, while for the other dataset, DT with PCA achieved the highest classification accuracy. Koushik Chowdhury et al. [29] conducted a comparative analysis of various classifiers, including NB, K-NN, LR, Gradient Boosting, SVM, DT, and Multi-Layer Perceptron(MLP) Classifier, to predict ASD. They assessed performance using evaluation metrics such as Accuracy, AUC Value, Precision, Recall, and F1 Score. Their results indicated that the MLP Classifier achieved the highest accuracy as compared to NB. Notably, SVM was identified as the most suitable classifier for the experimental dataset. Tania Akter et al. [30] investigated a variety of classifiers Extreme Gradient Boosting (XGB), DT, NB, RF, K-NN, Gradient Boost (GB), MLP, SVM, and LR on male and female datasets for predicting ASD across different age groups. They amalgamated datasets from toddlers, children, adolescents, and adults, preprocessed the data, and categorized it by gender. Their results indicated that all-female datasets exhibited higher positive efficacy than all-male datasets. Bhawana Tyagi et al. [31] assessed the effectiveness of various techniques such as K-NN, Linear Regression, Linear Discriminant Analysis(LDA), SVM, NB, and Classification and Regression Tree (CART) and developed a mobile application for ASD identification. Their conclusion indicated that the LDA algorithm yielded the best outcome when compared to other algorithms. Vaibhav Vishal et al. [32] evaluated various machine learning algorithms, including K-NN, LR, SVM, and NB, and analyzed these algorithms to identify specific traits associated with ASD. Their experimental data demonstrated that among all algorithms, the NB algorithm achieved a higher accuracy. Oh et al. [33] proposed a unique classification method based on EEG signals with nolinear features that are selected using the t-test. The features used for the experiment include approximate entropy, Kolmogorov complexity, fuzzy entropy, Kolmogorov-Sinai entropy, modified multiscale entropy, Tsallis entropy, wavelet entropy, Signal activity, Bispectrum, Cumulant, Hjorth. Sharma et al. [34] analyzed multiple ML models like K-NN, LMT, RF, NB, LDA and DTs for autism. The SVM exhibited the highest efficiency among all machine learning models. Sankar Ganesh Karuppasamy et al. [35] developed a convolutional neural network(CNN) model to identify ASD using brain patterns. Erik Linstead et al. [38] applied neural networks to forecast autism in individuals receiving behavioral therapy. The study tried to identify the connection between various factors like treatment intensity, supervision hours, age, and gender. Wiratsin et al. [39] presented an algorithm to evaluate factors that have more impact on the identification of ASD for various age groups. The study utilized ASD screening datasets containing general and behavioral development information, analyzed using the apriori algorithm, Chi-Square test, and Mutual information test. Eslami et al. [40] conducted research using techniques like SVM and RF for classifying and identifying ADHD and ASD. with the help of. They conducted the study using MRI data analysis to classify ASD and ADHD. Ahmed et al. [41] proposed a model that combines the Restricted Boltzmann Machine (RBM) and SVM algorithms for feature extraction from fMRI data. Grid-search cross-validation results demonstrated the framework's effectiveness in identifying ASD meltdowns. Patwary et al. [42] introduced a mathematical model based on fuzziness. The study explained the effect of low-fuzziness samples on ASD model learning performance. They investigated and identified that by incorporating low-fuzziness patterns into validation samples with labeled data semi-supervised machine learning and ANN techniques have more likelihood for early ASD diagnosis.

2.3. Key features used in machine learning models for autism detection

ML models for autism discovery commonly rely on a collection of essential attributes integral to their operation. These features largely relate to aspects associated with communication skills and social communications. They encompass parameters such as gaze fixation, non-verbal cues, facial expressions, vocal tonality, and linguistic aptitude, repetitive and restrictive behaviors, recurrent physical gestures. Sensory sensitivities, exemplified by heightened or diminished responsiveness to specific stimuli like sound, tactile sensations, or luminosity, are frequently integrated as features within ML models for autism detection. Cognitive aptitudes, comprising intelligence quotient (IQ) and specific cognitive profiles, are also taken into account in select models. Moreover, certain models incorporate familial medical history and genetic markers as features. This is used to discover genetic associations with autism which contributes to forecasting the possibility of autism indicators in individuals. Advanced models may further be able to capture distinctive patterns of brain activity by investigating the data from electroencephalography (EEG) or neuroimaging. Additionally, ML models may identify features extracted from questionnaires and diagnostic criteria that identify significant relations between the individual's behavioral traits.

3. Dataset and Preprocessing

3.1. Source of Data:

This investigation is carried out using binary (10 behavioral traits [44]) and categorical attributes. Table 1 shows selected binary attributes with abbreviations.

Table 1: Selected behavioral traits

Abbr.	Attribute
A1	Does your child look at you when you call his/her name?
A2	How easy is it for you to get eye contact with your child?
A3	Does your child point to indicate that s/he wants something? (e.g. a toy that is out of reach)
A4	Does your child point to share interest with you?!(e.g. pointing at an interesting sight)
A5	Does your child pretend? (e.g. care for dolls, talk on a toy phone)
A6	Does your child follow where you're looking?
A7	If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging them)
A8	Would you describe your child's first words as
A9	Does your child use simple gestures? (e.g. wave goodbye)
A10	Does your child stare at nothing with no apparent purpose?

Table 2 depicts the categorial attributes

Table 2: Categorial attributes

Abbr.	Attribute
Age	Age
Sex	Sex
Eth	Ethnicity
Jaundice	Born with jaundice

A dataset can be generated by utilizing a mobile app or website to examine autism in children. The identity is not revealed by any means. Websites like Kaggle.com, openml.org, etc have many datasets available. These datasets are used as test datasets and the details are given in Table 3.

Table 3: Dataset Details

Source	Total Features	No of Records
Kaggle.com [46]	16	1054
Kaggle.com [47]	16	797
Total Records		1851

Table 3 also illustrates the number of attributes and the total number of recorded inputs.

3.2. Data Encoding

The algorithms used in this paper have been implemented using the open-source programming language Python, which is freely accessible for download at https://www.python.org/. Python is one of the most widely adopted programming languages for ML and data science. It boasts a robust ecosystem replete with specialized libraries and frameworks tailored for the intricacies of ML, such as NumPy, Pandas, scikit-learn, TensorFlow, and PyTorch. These libraries are equipped with a variety of functions and tools crucial for tasks related to data manipulation, statistical analysis, and managing ML models. The versatility of Python's is in its support for a diverse spectrum of ML techniques. Furthermore, Python has the capability to construct machine learning workflows, and carry out data preprocessing and feature engineering that supports the phases of model training and evaluation.

Table 4: Numerical Encoding of Binary Attributes

Value / Attribute	0	0	1	1	1	..
A1	Always	Usually	Sometimes	Rarely	Never	..
A2	Always	Usually	Sometimes	Rarely	Never	..
A3	Always	Usually	Sometimes	Rarely	Never	..
A4	Always	Usually	Sometimes	Rarely	Never	..
A5	Always	Usually	Sometimes	Rarely	Never	..
A6	Always	Usually	Sometimes	Rarely	Never	..
A7	Always	Usually	Sometimes	Rarely	Never	..
A8	Always	Usually	Sometimes	Rarely	Never	..
A9	Always	Usually	Sometimes	Rarely	Never	..
A10	Rarely	Never	-	Always	Usually	..

Table 5: Numerical Encoding of categorial Attributes

Value Attribute	0	1	2	3	..
Ethnicity	African	Asian	Black	Hispanic	..
Jaundice	No	Yes	-	-	..
Sex	Male	Female	-	-	..

Table 4 and 5 illustrates attributes and their corresponding numerical encodings. A database of 1851 encoded individual's records is used for the ASD trait prediction. Table 6 illustrates the encoded database.

Table 6: Database with encoded values

Case No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Age	Sex	Ethnicity	Jaundice
1	0	0	0	0	0	0	1	1	0	1	28	1	5	1
2	1	1	0	0	0	1	1	0	0	0	36	0	10	1
3	1	0	0	0	0	0	1	1	0	1	36	0	5	1
4	1	1	1	1	1	1	1	1	1	1	24	0	3	0
5	1	1	0	1	1	1	1	1	1	1	20	1	10	0
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1851	0	0	0	0	0	0	1	0	0	1	36	0	7	0

3.3. Preprocessing steps

Preprocessing steps are essential for preparing raw data for further analysis or model building. Data Cleaning step involves removing or handling missing or incorrect data values, including Missing Data, Outliers, Duplicates etc. Data Transformation step involves transforming data into a more suitable form for analysis or modeling, including Normalization, One-Hot Encoding, Ordinal Encoder, Target Encoder, Feature Encoding, Variable Discretization, Feature Scaling. Feature Extraction/Selection step involves selecting or extracting relevant features that have the most predictive power to the problem being solved, including Dimensionality Reduction, Feature Engineering. These preprocessing steps are crucial for improving the data suitability and excellence for analysis or building ML models, leading to better accuracy and robustness in the final results.

4. Machine Learning Algorithms For Autism Prediction

ML algorithms are broadly used for anticipating ASD in various research studies. These algorithms utilize inherent data patterns and connections to recognize individuals with a potential risk of ASD by considering their traits, behaviors, and other pertinent factors. This study is focused on following machine-learning algorithms for ASD prediction.

4.1. Random Forests:

Random forests represent an ensemble learning technique that amalgamates numerous DTs to formulate predictive outcomes. This algorithm generates a large number of DTs and compiles their predictions to produce a final result. Prediction using RF can handle high-dimensional data and are known for their ability to handle noise and missing values. They have been utilized to predict ASD by considering a diverse range of attributes and achieving high

$$f(a,i) = \begin{cases} 1 \text{ if } a_i = 1 \\ 0 \text{ if } a_i = 0 \end{cases} \quad (1)$$

Function $f(a,i)$ is used for finding the outcome for an attribute i which varies from 1,2,…n. The final score p can be derived as

$$\boldsymbol{p} = \sum_{i=1}^{n} \boldsymbol{f(a,i)} \quad (2)$$

Pseudo Code: RFASDPrediction (D, c)
Input: D: ASD database, c: sets of applicable rules;
Output: R: Sets of results.
Begin
Step 1: Call train, test = DataPreprocessing(D,c)
Step 2: Build individual DT for samples with a subset of data and a subset of features.
Step 3: Calculate class labels by Majority Voting or averaging
Step 4: Return the result R
Stop

Pseudo Code: DataPreprocessing(D,c)
Input: D: ASD database, c: sets of applicable rules;
Output: train: training data set, test: testing dataset
Begin
Step 1: Prepare train as n records from dataset D
Step 2: Prepare test as remaining records from dataset D
Step 3: return train, test
Stop

Labelled dataset, where the traits and corresponding labels are provided, is passed as input to algorithm. The database is divided into training (TN) and testing (TS) sets using the DataPreprocessing function. The algorithm randomly selects a subset of the TN to create a new dataset for each DT. A DT is constructed for each subset using a randomly picked collection of features. Each tree's prediction is considered as a vote to assign class labels to test data and majority vote is considered for final predictions.

4.2. Naive Bayes:

In ASD forecast, Naive Bayes has been applied to gauge the likelihood of an individual having ASD by evaluating the probability of various features being present. NB applies Bayesian probability principles to classify data. The assumption of this algorithm is that the presence of a specific attribute within a category is independent of the existence of other features.

According to Bayes theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

where:

P(A|B) : Restrictive Probability of A given B
P(B|A) : Restrictive Probability of B given A
P(A) : Probability of A
P(B) : Probability of B
$f(a,i) = P(a_i)$

Function $f(a,i)$ is the function used for finding the probability for an attribute i which varies from 1,2,…n that contribute to ASD. The final score ASDscore can be calculated as score

ASDscore = P(S=n)

Here S is the sum of the n attribute variables.

Pseudo Code: NASDPrediction (D, c)
Input: D: ASD database, c: sets of applicable rules;
Output: R: result set.
Begin
Step 1: Call train, test = DataPreprocessing(A,c)
Step 2: Calculates the prior probabilities of each class.
Step 3: Call P = Extract_Features_Likelyhood_Calc(train)
Step 4: R = Call Predict(P, test)
Step 5: Return the result R
Stop

Pseudo Code: DataPreprocessing(D,c)
Input: D: ASD database, c: sets of applicable rules;
Output: train: training data set, test: testing dataset
Begin
Step 1: Prepare train as n records from dataset D
Step 2: Prepare test as remaining records from dataset D
Step 3: return train , test
Stop

Pseudo Code: Extract_Features_Likelyhood_Calc(train)
Input: train: Training set
Output : P: Probabilities
Begin
Step 1: Extract the relevant features form training set.
Step 2: Compute likelihood of each feature.
Step 3: Estimate P by multiplying prior probability with likelihood probability for each feature.
Step 4: Return the result P
Stop

Pseudo Code: Predict(P, test)
Input : P: test: test dataset
Output : R : Result Set
Begin
Step 1: Calculate posterior probability by multiplying the prior and likelihood probabilities
Step 2: Class label are assigned by considering the peak probability.
Step 3: Return Result R
Stop

The algorithm necessitates an annotated dataset as its primary input, encompassing both feature vectors and their corresponding class labels. Subsequently, the dataset is bifurcated into training and testing subsets using the DataPreprocessing function. Following this initial step, the algorithm calculates the prior probabilities of each class by tallying the occurrences of individual class labels within the training subset. The function, Extract_Features_Likelihood_Calc extracts features for the likelihood calculation of each feature by counting the occurrences of features for each class. The algorithm then calculates the product of prior and likelihood probabilities for each individual feature, within the context of each class. The Predict function is invoked to predict the unclassified data. A new instance with the unknown class label is given as input. Posterior probabilities for each class are calculated and a class label with the highest probability is predicted as output.

4.3. Support Vector Machines (SVM):

SVM is a classification algorithm that separates data points using hyperplanes in high-dimensional spaces. Maximizing the margin between different classes while minimizing the classification error is the aim of this algorithm. To predict autism the input data is transformed into a higher-dimensional space and the optimal hyperplane is identified that can differentiate between ASD and non-ASD individuals.

Pseudo Code: SASDPrediction (D, c)
Input: D: ASD database, c: sets of applicable rules;
Output: R: result set.
Begin
Step 1: Call train, test = DataPreprocessing(A,c)
Step 2: Find separating hyperplanes used to classify train data into ASD and NonASD classes.
Step 3: Identifies the best hyperplane
Step 4: Predict the outcome for test data.
Step 5: Return the result R
Stop

Pseudo Code: DataPreprocessing(D,c)
Input: D: ASD database, c: sets of applicable rules;
Output: train: training data set, test: testing dataset

Begin
Step 1: Prepare train as n records from dataset D
Step 2: Prepare test as remaining records from dataset D
Step 3: return train, test
Stop

Given a labeled training dataset with various data points, the algorithm aims to find the best possible decision boundary that separates the data into different classes or predicts the output value for regression. Using the DataPreprocessing function, the dataset is separated into training and testing sets. The algorithm uses a technique called the kernel trick to convert the input data into a higher-dimensional feature space. It then identifies the best hyperplane, which is the decision boundary that maximizes the margin between the classes or regression output. The algorithm checks on which side of the decision boundary test data points fall to make the predictions. If they fall on one side, they are classified as belonging to one class(ASD trait); if they fall on the other side, they belong to the other class(nonASD trait).

5. Results and Analysis

Three distinct algorithms have been created and put into practice using Python. The first one employs an ensemble learning approach and is called the RASDPrediction algorithm. The second is the NASDPrediction algorithm, which utilizes a probabilistic classifier. The third, SASDPrediction algorithm, is built on statistical learning theory and the principle of maximum margin classification. The results obtained by these algorithms are shown in Table 7 in the form of confusion matrix.

Table 7 : Confusion Matrix

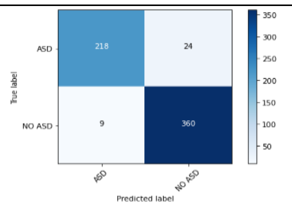
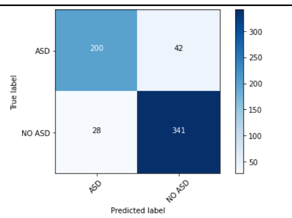
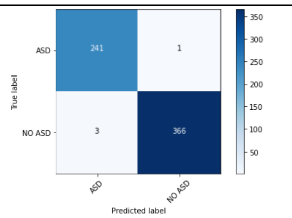
Algorithm	Confusion Matrix	AUC-ROC
RASDPrediction		0.94
NASDPrediction		0.88
SASDPrediction		0.99

Table 8 gives details of precision, recall, F1 score and accuracy obtained by these algorithms. F1-score is a statistical measure of a test's accuracy. It is calculated as the harmonic mean of the test's precision and recall. In other words, it balances the test's ability to correctly identify positive cases (precision) with its ability to correctly identify negative cases (recall).

Table 8: Analysis of Algorithms

Algorithm		Precision	Recall	F1 Score	Support
RASDPrediction	0	0.96	0.90	0.93	242
	1	0.94	0.98	0.96	369
	accuracy			0.95	611
	macro avg	0.95	0.94	0.94	611
	weighted avg	0.95	0.95	0.95	611
NASDPrediction	0	0.88	0.83	0.85	242
	1	0.89	0.92	0.91	369
	accuracy			0.89	611
	macro avg	0.88	0.88	0.88	611
	weighted avg	0.89	0.89	0.88	611
SASDPrediction	0	0.99	1.00	0.99	242
	1	1.0	0.99	0.99	369
	Accuracy			0.99	611
	Macro avg	0.99	0.99	0.99	611
	Weighted avg	0.99	0.99	0.99	611

6. Challenges and Limitations

One of the main challenges in predicting ASD with ML algorithms is the lack of standardized datasets. ASD is a complex disorder with heterogeneity in symptoms, making it difficult to collect and curate large-scale, high-quality datasets that represent the diverse characteristics of individuals with ASD. Due to the rarity of ASD, obtaining a large and representative sample size for training and testing machine learning algorithms can be challenging. Limited sample size can lead to overfitting or lack of generalizability of the trained models, resulting in low predictive accuracy. Handling imbalanced datasets is one of the reasons for the poor prediction of ASD. It requires careful handling of data imbalance techniques such as oversampling, under-sampling, or using specialized algorithms designed for imbalanced data. ASD is a developmental disorder, and symptoms may change over time. Predicting ASD accurately requires considering the temporal dynamics and capturing the changes in symptoms and behavior throughout development. Overall, addressing these challenges and limitations is necessary to develop effective and reliable machine learning algorithms for ASD prediction, aiding early prediction and diagnosis.

7. Future Directions

Future research areas may call for more analysis and improvement of ML algorithms for ASD identification and diagnosis. Overall, this study highlights the potential of machine learning in autism prediction. It offers a promising avenue for improving the efficiency and reliability of ASD assessment, with the potential to impact early intervention strategies and ultimately improve the lives of individuals with ASD.

8. Conclusion

Machine learning algorithms can be used to ASD based on Q Chat 10 questionnaire. The findings revealed that these algorithms can accurately predict whether an individual has ASD or not with a high degree of success. The implications of this research are promising for autism detection and diagnosis. Traditional methods of diagnosing ASD which can be time-consuming and prone to error, heavily rely on behavioral observation and subjective assessment. However, by leveraging machine learning, clinicians may have a more objective and efficient tool for early diagnosis of ASD, allowing for earlier intervention and improved outcomes for individuals with ASD.

References

[1] Maenner, M.J., Shaw, K.A., Bakian, A.V., Bilder, D.A., Durkin, M.S., Esler, A., Furnier, S.M., Hallas, L., Hall-Lande, J., Hudson, A. and Hughes, M.M., 2021. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveillance Summaries*, *70*(11), p.1.

[2] American Academy of Pediatrics. (2021). Identification and Evaluation of Children With Autism Spectrum Disorders. Pediatrics, 145(1), e20203447. https://doi.org/10.1542/peds.2020-3447

[3] Mazurek MO, Handen BL, Wodka EL, Nowinski L, Butter E, Engelhardt CR. Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and cost-effectiveness. JAMA Pediatr. 2020;174(6):e196333. doi:10.1001/jamapediatrics.2019.6333

[4] Anbarasan D, Vasudevan V. Autism spectrum disorder diagnosis using artificial intelligence models: A review. J Med Syst. 2020;44(3):69. doi:10.1007/s10916-020-1532-8

[5] Krishnaveni R, Pandian PS. Autism detection using an optimal deep learning approach. Bioinformation. 2019;15(2):110-114. doi:10.6026/973206300150110

[6] Chaidee N, Uyyanonvara B. Deep neural networks for autism spectrum disorder prediction using eye-tracking data. Appl Sci. 2020;10(13):4507. doi:10.3390/app10134507

[7] Varma P, Zhang X, Das R, Hussain M, Sundstedt V. Predicting autism severity from fMRI data using deep learning models. Sci Rep. 2020;10(1):1551. doi:10.103f8/s41598-020-58477-2

[8] Bieleninik Ł, Ghosh R, Howlin P, et al. Enhanced Autism Detection System (EADS) using machine learning models. PLoS One. 2019;14(12):e0225149. doi:10.1371/journal.pone.0225149

[9] Aoki Y, Abe T, Nishida K, et al. Predicting social impairment severity from facial emotion perception using deep learning in individuals with autism spectrum disorder. Sci Rep. 2020;10(1):12343. doi:10.1038/s41598-020-69290-w

[10] American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

[11] Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. Journal of autism and developmental disorders, 31(2), 131-144.

[12] Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the Modified Checklist for Autism in Low-Risk Toddlers. Pediatrics, 131(4), e1121-e1127.

[13] Maenner, M. J., Schieve, L. A., Rice, C. E., Cunniff, C., Giarelli, E., Kirby, R. S., ... & Yeargin-Allsopp, M. (2014). Frequency and pattern of documented diagnostic features and the age of autism identification. *Journal of the American Academy of Child & Adolescent Psychiatry, 53(6)*, 636-645.

[14] Robins, D. L., Casagrande, K., Barton, M. L., Chen, C. M., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F). Pediatrics, 133(1), 37-45.

[15] Dumont-Mathieu, T., Fein, D., Barton, M. L., & Robins, D. L. (2009). Monitoring autism spectrum disorders in public schools using the Modified Checklist for Autism in Toddlers (M-CHAT*). Journal of autism and developmental disorders, 39(6),* 873-880.

[16] Stone, W. L., Coonrod, E. E., Ousley, O. Y., & Hepburn, S. L. (2000). Screening tool for autism in two-year-olds (STAT): development and preliminary data. *Journal of autism and developmental disorders, 30(6)*, 607-612.

[17] Herlihy, L., Knoch, K., Vibert, B., & Fein, D. (2014). Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. Autism, 18(6), 728-737.

[18] Stenberg, N., Bresnahan, M., Gunnes, N., Hirtz, D., Hornig, M., Lie, K. K., ... & Thompson, W. K. (2014). Identifying children with autism spectrum disorder at 18 months in a general population sample. Paediatric and perinatal epidemiology, 28(3), 255-262.

[19] Sturner, R., Howard, B., Bergmann, P., Morrel, T., Andon, L., Hartmann, R., ... & Marks, D. (2016). Brief report: Development and replication of the screening tool for autism in two-year-olds (STAT). *Journal of autism and developmental disorders, 46(10),* 3619-3625.

[20] Tan, T., Hope, D., Rowe, P., & Reece, L. J. (2016). Early intervention services for infants and toddlers at risk for autism spectrum disorder in South Australia: Perspectives of parents and professionals. Research in Autism Spectrum Disorders, 27, 43-55.

[21] American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

[22] Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule – 2nd edition (ADOS-2). Torrance, CA: Western Psychological Services

[23] Chawarska, K., & Volkmar, F. R. (2007). Autism in infancy and early childhood. London, UK: Guilford Press.

[24] Lord, C., & Jones, R. M. (2012). Annual research review: Re-thinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 53(5),* 490-509.

[25] Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., ... & Fein, D. (2015). Early intervention for children with autism spectrum disorder under 3 years of age: recommendations for

practice and research. Pediatrics, 136(Supplement 1), S60-S81.

[26] Di Martino, A., Yan, C. G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., ... & Wolfers, T. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry, 19(6), 659-667.

[27] Courchesne, E., Mouton, P. R., Calhoun, M. E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M. J., & Barnes, C. C. (2011). Neuron number and size in prefrontal cortex of children with autism. *JAMA, 306(18),* 2001-2010.

[28] Lohar, M. and Chorage, S., 2021, October. Automatic Classification of Autism Spectrum Disorder (ASD) from Brain MR Images Based on Feature Optimization and Machine Learning. In *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)* (pp. 1-7). IEEE.

[29] Chowdhury, K. and Iraj, M.A., 2020, November. Predicting autism spectrum disorder using machine learning classifiers. In *2020 international conference on recent trends on electronics, information, communication & technology (RTEICT)* (pp. 324-327). IEEE

[30] Akter, T. and Ali, M.H., 2021, December. Predicting Autism Spectrum Disorder Based On Gender Using Machine Learning Techniques. In *2021 3rd International Conference on Electrical & Electronic Engineering (ICEEE)* (pp. 185-188). IEEE.

[31] Tyagi, B., Mishra, R. and Bajpai, N., 2018, November. Machine Learning Techniques to Predict Autism Spectrum Disorder. In *2018 ieee punecon* (pp. 1-5). IEEE.

[32] Vishal, V., Singh, A., Jinila, Y.B., Shyry, S.P. and Jabez, J., 2022, April. A Comparative Analysis of Prediction of Autism Spectrum Disorder (ASD) using Machine Learning. In *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 1355-1358). IEEE.

[33] Oh, S.L., Jahmunah, V., Arunkumar, N., Abdulhay, E.W., Gururajan, R., Adib, N., Ciaccio, E.J., Cheong, K.H. and Acharya, U.R., 2021. A novel automated autism spectrum disorder detection system. *Complex & Intelligent Systems*, *7*(5), pp.2399-2413.

[34] Sharma, A. and Tanwar, P., 2020, June. Deep analysis of autism spectrum disorder detection techniques. In *2020 International Conference on Intelligent Engineering and Management (ICIEM)* (pp. 455-459). IEEE.

[35] Karuppasamy, S.G., Muralitharan, D., Gowr, S., Arumugam, S.R., Devi, E.A. and Maharajan, K., 2022, April. Prediction of Autism Spectrum Disorder using Convolution Neural Network. In *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 1096-1100). IEEE.

[36] Yang, X., Sarraf, S. and Zhang, N., 2018. Deep learning-based framework for Autism functional MRI image classification. *Journal of the Arkansas Academy of Science, 72(1),* pp.47-52.

[37] Saleh, A.Y. and Chern, L.H., 2021. Autism Spectrum Disorder Classification Using Deep Learning. *International Journal of Online & Biomedical Engineering*, *17(8).*

[38] Linstead, E., German, R., Dixon, D., Granpeesheh, D., Novack, M. and Powell, A., 2015, December. An application of neural networks to predicting mastery of learning outcomes in the treatment of autism spectrum disorder. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 414-418). IEEE.

[39] Wiratsin, I.O. and Narupiyakul, L., 2021, January. Feature selection technique for autism spectrum disorder. In *Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence* (pp. 53-56).

[40] Eslami, T., Almuqhim, F., Raiker, J.S. and Saeed, F., 2021. Machine learning methods for diagnosing autism spectrum disorder and attention-deficit/hyperactivity disorder using functional and structural MRI: a survey. *Frontiers in neuroinformatics*, *14*, p.575999.

[41] Ahmed, M.R., Ahammed, M.S., Niu, S. and Zhang, Y., 2020, March. Deep learning approached features for ASD classification using SVM. In *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)* (pp. 287-290). IEEE.

[42] Patway, M.J. and Wang, X.Z., 2019. Sensitivity analysis on initial classifier accuracy in fuzziness based semi-supervised learning. *Information Sciences*, *490*, pp.93-112.

[43] Patwary, M.J., Wang, X.Z. and Yan, D., 2019. Impact of fuzziness measures on the performance of semi-supervised learning. *International Journal of Fuzzy Systems, 21(5),* pp.1430-1442.

[44]. https://www.autismalert.orguploadsPDFSCREENING--AUTISM--QCHAT-10%20Question%20Autism%20Survey%20for%20Toddlers.pdf

[45] C., Baron-Cohen, S., Wheelwright, S., Charman, T., Richler, J., Pasco, G., & Brayne, C. (2008). The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): a normally distributed quantitative measure of autistic traits at 18–24 months of age: preliminary report. *Journal of autism and developmental disorders*, *38*, 1414-1425.

[46] Fadi, “Autism screening for toddlers,” Kaggle, 2018. [Online]. Available: https://www.kaggle.com/fabdelja/autism-screening-for-toddlers. [Accessed: 09-July-2023].

[47] Basma Elshoky, “Autism screening child two version,” Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/basmarg/autism-screening-child-two-version. [Accessed: 09-July-2023].

[48] http://www.autism-india.org/about-autism.php

[49] Rajendra B Patil, B V Pawar and Ajay S Patil. Article: Sperm Donor Selection using Nominal and Binary Variable Methods. *International Journal of Computer Applications 130(14):*20-24, November 2015. Published by Foundation of Computer Science (FCS), NY, USA

[50] Yenduri, G., Kaluri, R., Rajput, D. S., Lakshmanna, K., Gadekallu, T. R., Mahmud, M., & Brown, D. J. (2023). From Assistive Technologies to Metaverse–Technologies in

Int. J. Advanced Networking and Applications
Volume: 15 Issue: 04 Pages:100-200 (2023) ISSN: 0975-0290 (Special Issue – ASCIS 2023)

Inclusive Higher Education for Students with Specific Learning Difficulties: A Review. *IEEE Access*.

Biographies and Photographs

Ms. Pournima P Bhangale has over two decades of professional experience in education and learning field, for bachelor and master studies in Information Technology. It is with high commitment with focused on accomplishing success, currently perusing doctorate Information Technology. The central area of interest and proficiency lies in Programming Languages (C / C++ / JAVA / Python / R) , Artificial Intelligence, and Data Analytics.

Dr. Rajendra Patil is a dedicated and result-oriented professional with 2 decades of experience in the area of research, training, and consultancy. A proven track record of successful implementation of various research and consultancy projects using innovative methods. He is very passionate about training and product delivery for new innovative technology. He has 20+ years of experience in the Learning and Development domain performing various roles as a Faculty, Author, Content Writer Corporate Trainer, IT and Academic Operations, Researcher, Technical Consultant, Advisor, and Subject Matter Expert. Key interest and expertise in Programming Languages, Data Analytics, and Business intelligence, published white papers. Delivered 100+ sessions on Python R, data mining, and data warehousing. Published 40 Journal Papers and 5 Reference books on R Programming, Data Science, PL/SQL, and Business Intelligence. He is a corporate trainer for Programming and other Product-based training like HIMS. He is a passionate researcher of data science/AI-ML.