Department of Computer Science and Engineering

University of Barishal

PAPERS SUMMARY OF

Machine Learning for Autism Spectrum Disorder Classification and Assessment: A Holistic Perspective on Current Trends

---

Submitted By

Rupa Samodder

Computer Science and Engineering

Roll: 19CSE011

Submitted To

Dr. Tania Islam

Assistant professor

Department of Computer Science and Engineering

University of Barisal

November 09, 2023

Abstract

This paper presents a summary of **five research papers** focusing on the application of machine learning in diagnosing and assessing Autism Spectrum Disorder (ASD). These papers address various aspects of ASD, including classification in adults and children, psychological assessment in school and community settings, and optimized classification methods. Together, these papers offer a comprehensive overview of how machine learning is advancing the understanding and diagnosis of ASD.

Title of the papers are:

1. Classification of adult autistic spectrum disorder using machine learning approach.
2. Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques.
3. Toward Machine Learning-Based Psychological Assessment of Autism Spectrum Disorders in School and Community.
4. Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques.
5. Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder.

Contents

List of Figures

List of Tables

1 Introduction

Autism Spectrum Disorder or ASD, is a neurological disease that affects children and adults alike [1]. It is generally characterized by speech and language impairments, mutual communication difficulties, and limited activities, identified in the first two years of life [2].

This study "**Classification of adult autistic spectrum disorder using machine learning approach**" focuses on using machine learning techniques for early diagnosis of Autism Spectrum Disorder (ASD) based on a dataset with 16 selected attributes and 703 patients and non-patients. Various methods, including **Linear support vector machine (SVM), k-nearest neighbours (k-NN), J48, Bagging, Stacking, AdaBoost,** and **naïve bayes**, were employed with 3, 5, and 10-folds cross-validation to achieve high accuracy in predicting ASD status by the Waikato environment for knowledge analysis (WEKA) platform, emphasizing the importance of early detection for this lifelong condition [1].

Autism Spectrum Disorder (ASD), impacting around 1% of the population globally, often emerge in early childhood due to genetic or environmental factors [3]. To expedite diagnosis and improve accuracy, machine learning methods, including **Support Vector Machines (SVM), Random Forest Classifer (RFC), Naïve Bayes (NB), Logistic Regression (LR),** and **KNN** are applied to complement clinical tests, addressing the need for timely identification and access to therapies in "**Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques**" paper [3]. Screening tools like Q-CHAT-10 [4] for toddlers are employed to enhance early detection.

Between 2009 and 2017, a report from the Centers for Disease Control and Prevention indicates, 17% of children between the ages of three and seventeen received diagnoses of developmental disabilities [5]. **"Toward Machine Learning-Based Psychological Assessment of Autism Spectrum Disorders in School and Community"** study introduces a questionnaire-based method, developed from previous research, and evaluates various machine learning classifiers. Notably, artificial neural networks achieved the highest accuracy at 89.8%, suggesting its potential as a screening tool for ASD detection [6].

"**Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques**" paper evaluates the performance of **Naïve Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network,** and **Convolutional Neural Network** for predicting and analyzing ASD problems in children, adolescents, and adults [2]. The results show that CNN-based prediction models work better on all three datasets, with higher accuracy of 99.53%, 98.30%, and 96.88% for ASD screening in adults, children, and adolescents, respectively. This paper suggests that CNN-based models are a promising tool for ASD detection [2].

The study "**Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder**" focuses on ASD detection, evaluating machine learning algorithms, their effectiveness, and pre-processing techniques for medical datasets related to predicting early autism traits in toddlers and adults [7]. It emphasizes the need for faster diagnosis in healthcare, particularly in ASD cases, where it currently takes up to 6 months [7].

2 Review of Literature

Numerous recent research studies have harnessed machine learning techniques in diverse ways to enhance and expedite the detection of Autism Spectrum Disorder. Machine learning methods used in "**Classification of adult autistic spectrum disorder using machine learning approach**" study can substantially contribute new methods to diagnosis cases related to ASD. M. Duda, R. Ma et al.[8] used data mining techniques, such as decision trees, random forests, SVM, logistic regression, categorical lasso, and LDA, were used to evaluate AUC-based classification accuracy on a dataset of 2775 autism subjects from Simons Simplex Collection, Boston Autism Consortium, and Autism Genetic Resource Exchange, resulting in a 96.5% accuracy with SVM after feature selection. A dataset for ASD screening in children, comprising 292 subjects, with 141 diagnosed patients, was employed to diagnose ASD disease. LDA outperformed k-NN with an accuracy of 90.80% [9].

To provide a concise view of literature survey, in the paper "**Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques**" there is a table of the most relevant paper summarizations. Thabtah et al. [10] introduced a novel machine learning method known as Rules-Based Machine Learning (RML), which not only identifies ASD traits but also provides users with a set of rules to gain insights into the factors contributing to the classification. Using the ABIDE database, Li et al. [11] derived six individual attributes from a sample of 851 subjects and applied a cross-validation approach to train and test machine learning models, enabling the classification of individuals with and without ASD.

Apart from the conventional machine learning approaches "**Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques**" explored the possibility of applying deep learning based models. Vaishali R, Sasikala R. et al. [12] proposed a method for Autism identification using optimal behavior attributes from a 21-feature ASD diagnosis dataset, employing swarm intelligence-based binary firefly feature selection to demonstrate that only 10 features are required for distinguishing between ASD and non-ASD patients, achieving an average accuracy between 92.12% and 97.95%. J. A. Kosmicki1 et al. [13] applied machine learning to assess ASD based on the ADOS subset of children's behaviors, utilizing eight different algorithms and feature selection to identify ASD risk, achieving 98.27% and 97.66% accuracy using 9 and 12 selected behaviors.

In paper **"Toward Machine Learning-Based Psychological Assessment of Autism Spectrum Disorders in School and Community"** have conducted a survey in schools and communities leveraging the questionnaire and prepared a dataset. Here Hossain et al.[14] assessed 25 machine learning classifiers in an ASD dataset, determining that SVM with Sequential minimal optimization outperformed others, highlighting the challenges in cataloging various physiological and psychological aspects by health professionals.

3 Methodology

3.1 Proposed Methods

This section outlines the methodological approach to improve existing machine learning algorithms for Autism Spectrum Disorder (ASD) detection, based on the limitations identified in the current research. The methodology focuses on dataset preparation, algorithm enhancements, and performance evaluation. Figure 1. shows the steps in the proposed workflow which involves the pre-processing of data, training, and testing with specified models, evaluation of results and prediction of ASD.

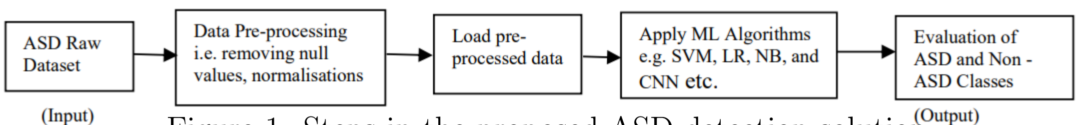

Figure 1: Steps in the proposed ASD detection solution

1. **Data Collection and Preprocessing:** The research will utilize the publicly available ASD datasets, such as the UCI Machine Learning Repository ASD dataset for adults, children, and adolescents. The primary features from these datasets include demographic details (age, gender, etc.), medical history (jaundice at birth), behavioral screening results (AQ-10 test for adults), and ASD traits.

2. **Algorithm Enhancements:** The study aims to run the performance of basic machine learning algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Trees, Random Forests for better accuracy.

3. **Cross-Validation and Hyperparameter Tuning:** To ensure robustness and avoid overfitting, k-fold cross-validation will be employed, with k = 5 or 10, depending on dataset size. This will provide a reliable estimate of model performance across different subsets of data.

4. **Evaluation of classes:** Performance will be evaluated on accuracy, precision and recall, F1-Score, ROC-AUC, Confusion Matrix to further refine the models.

3.2 Data Description and Preprocessing

The data for adult ASD screening was sourced from the UCI machine learning repository [15], using the ASD test mobile application by Thabtah [16] includes 703 subjects with 21 features, categorizing adults into those with ASD (189 subjects) and those without ASD (515 subjects). In the ASD adult dataset, missing values were substituted, and irrelevant attributes like ethnicity, country of residence, app usage, age description, and relation will be omitted to improve classification accuracy. The final dataset will be reduced to 16 essential features, including age, gender, jaundice, autism, screening score, 1-10 autism-related behavioral questions, and class/ASD. Numerical features will be transformed into nominal attributes using discretization.

Features	Type	Description
Age	Number	The age of the subjects
Gender	String	The individuality can be female or male
Ethnicity	String	The ethnicity of the subject
Jaundice	Boolean (yes or no)	If the case was diagnosed with jaundice
Autism	Boolean (yes or no)	If the close relatives have PDD
Relation	String	The person who completed the test such as the individual, parents, caretakers, and physicians
Country of residence	String	The country residence of the subject
Used app before	Boolean (yes or no)	If the person has used the screening application
AQ-1	Binary (0, 1)	The response is clarified based on the screening process
AQ-2	Binary (0, 1)	The response is clarified based on the screening process
AQ-3	Binary (0, 1)	The response is clarified based on the screening process
AQ-4	Binary (0, 1)	The response is clarified based on the screening process
AQ-5	Binary (0, 1)	The response is clarified based on the screening process
AQ-6	Binary (0, 1)	The response is clarified based on the screening process
AQ-7	Binary (0, 1)	The response is clarified based on the screening process
AQ-8	Binary (0, 1)	The response is clarified based on the screening process
AQ-9	Binary (0, 1)	The response is clarified based on the screening process
AQ-10	Binary (0, 1)	The response is clarified based on the screening process
Age description	Text	Age category
Screening score	Integer	The total score was determined using the implementation of the screening algorithm
Class/ASD	Boolean (yes or no)	The result is shown after the test

Figure 2: Features and its descriptions

The dataset is divided into three category with multiple instanes.

Sr. No.	Dataset Name	Sources	Attribute Type	Number of Attributes	Number of Instances
1	ASD Screening Data for Adult	UCI Machine Learning Repository [12]	Categorical, continuous and binary	21	704
2	ASD Screening Data for Children	UCI Machine Learning Repository [15]	Categorical, continuous and binary	21	292
3	ASD Screening Data for Adolescent	UCI Machine Learning Repository [16]	Categorical, continuous and binary	21	104

Figure 3: List of ASD datasets

3.3 Classifier Algorithms

- **K-Nearest Neighbors (KNN):**

The K-Nearest Neighbors algorithm is a supervised machine learning method used for classification and regression[1]. It operates under the assumption that similar data points are located in closely, where 'K' signifies the number of seed points to be chosen, requiring a careful selection to minimize errors [2]. It relies on the concept of similarity through factors like distance or nearest neighbor identification[1]. By finding the k data points nearest to a new data point x using the Euclidean distance metric, KNN employs majority voting to assign a label to x, and values of k (k=1 to k=10) yielded the highest accuracy [3].

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

- **Naive Bayes (NB):**

Naïve bayes is one of the supervised machine learning approaches that are mainly known as Bayesian algorithms with a simple conditional probability distribution [1]. The main principle of naïve bayes is focused on the expectations of freedom, which indicates less training time to be compared to the SVM approach [1]. It calculates the posterior probability for a dataset using the prior probability and likelihood. [2].

$$P(c|x) = \frac{P(X|C)P(c)}{P(x)}$$

- **Support Vector Machine (SVM):**

SVM is a linear supervised machine learning approach that is used for classification and regression. It is a pattern recognition problem solver. It does not cause the problem

Paper Title	**Classification Algorithms**
Classification of adult autistic spectrum disorder using machine learning approach	KNN, SVM, NB, J48 decision tree, AdaBoost, Bagging, Stacking
Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques	KNN, SVM, NB, LR, RFC
Toward Machine Learning-Based Psychological Assessment of Autism Spectrum Disorders in School and Community	KNN, SVM, RFC, ANN
Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques	KNN, SVM, NB, LR, ANN, CNN
Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder	KNN, LR, RFC

Table 1: Used Classification Algorithms

of overfitting [2]. SVM separates the classes by defining a decision boundary aiming to maximize the margin between the decision hyperplane and the nearest data point [1].

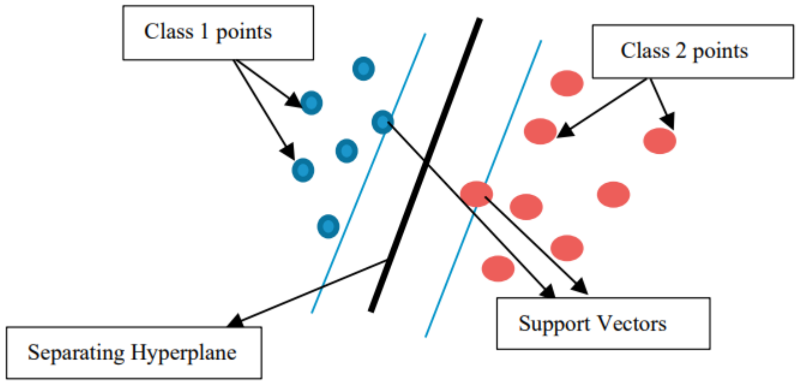

Figure 4: An SVM classifier

- **J48 decision tree:**

J48 decision tree is a machine learning approach [1]. Generally, J48 is used to develop a classification tree based on a hierarchical tree system, in which the decision results have illustrated the attributes and terminal nodes. The visual classification of the J48 approach is effective and efficient. Nevertheless, J48 is vulnerable to the noise in the data [1].

- **AdaBoost:**

AdaBoost is a supervised algorithm where core idea is to match a sequence of weak learner models that are more effective than random guessing[1]. Each instance in the training dataset is weighted to determine the accuracy either it is classified correctly or incorrectly. The decision stump is used to classify the AdaBoost models[1]. The primary purpose of the decision stump is to boost the AdaBoost M1 nominal classifier. Only minor class problems can be tackled. The final prediction is then obtained from the combination of the predicted model based on a weighted majority vote (classification) or weighted sum (regression)[1].

- **Bagging:**

Bagging is one of the most popular techniques in ensemble methods and is known as bootstrap aggregation [1]. This method can be used to reduce the variance for the algorithms that have high variance such as decision trees. The algorithm generates a decision tree and prunes it with a reduced-error with back fitting. The lack of values was coped with by dividing the corresponding instances into bits. The final decision tree was obtained as a composition of all base classifiers with the maximum votes [1].

- **Stacking:**

Stacking is an ensemble machine learning approach used to integrate either diversified classification or regression through meta-classifiers [1]. The features on the results of the base level are prepared using a proper training set that contains various machine learning approaches. Thus, stacking is a stratified approach [1].

- **Logistic Regression (LR):**

Logistic Regression aims to find the best-fitting model to describe the relationship between a binomial character of interest and a set of independent variables by utilizing a logistic function to fit the data points [3].

- **The Random Forest Classifier (RFC):**

The Random Forest Classifier (RFC) is a versatile algorithm for classification, regression, and other tasks, employing multiple decision trees on random data points and selecting the best solution by voting [3].

- **Artificial Neural Network (ANN):**

A deep learning model with one input layer, three fully connected hidden layers, two batch normalization layers, two dropout layers, and an output layer [6].

4 Results and Discussion

4.1 Evaluation Matrix

Usually, in most predictive models, the data points lie in the following four categories:

- **True positive (TP):** The individual has ASD and we predicted correctly that the individual has ASD.
- **True negative (TN):** The individual does not have ASD and we predicted correctly that the individual does not have ASD.
- **False positive (FP):** The individual does not have ASD, but we predicted incorrectly that the individual has ASD. This is known as Type 1 error.
- **False negative (FN):**The individual has ASD, but we predicted incorrectly that the individual does not have ASD. This is known as Type 2 error.

The above four categories when put together in the form of a matrix produce the confusion matrix. The confusion matrix is particularly useful in gauging the performance of a machine learning classification model [3]. The performance generally measured by accuracy, precision, recall, and F1 score.

- **Accuracy** is the simplest metric, and it measures the percentage of predictions that the model makes correctly [3]. For example, if a model predicts that 100 examples are positive and 90 of those predictions are correct, then the model has an accuracy of 90% .

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}.$$

- **Precision** measures the percentage of positive predictions that are actually correct [3]. For example, if a model predicts that 100 examples are positive and 90 of those predictions are correct, then the model has a precision of 90%.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

- **Recall** measures the percentage of actual positives that the model correctly predicts [1]. For example, if there are 100 actual positive examples and the model predicts that 90 of them are positive, then the model has a recall of 90%.
- **F1 score** is a harmonic mean of precision and recall, and it is often used to evaluate the performance of machine learning models when there is a trade-off between precision and recall [1]. For example, if a model has a precision of 90% and a recall of 80%, then the model has an F1 score of 85.7%.

$$\text{Recall} = \frac{TP}{TP + FN}.$$

4.2 Result and Comparison of Classifiers

"**Classification of adult autistic spectrum disorder using machine learning approach**" study found that several machine learning approaches, including linear SVM, naïve Bayes, J48, bagging, and stacking, achieved 100% classification accuracy for ASD data [1]. These approaches outperformed other methods, such as k-NN, which achieved 99.1% accuracy. The study also found that 10-fold cross validation was more effective than 3-fold cross validation in improving classification accuracy. K-NN and Stacking showed performance improvements with an increased number of folds [1].

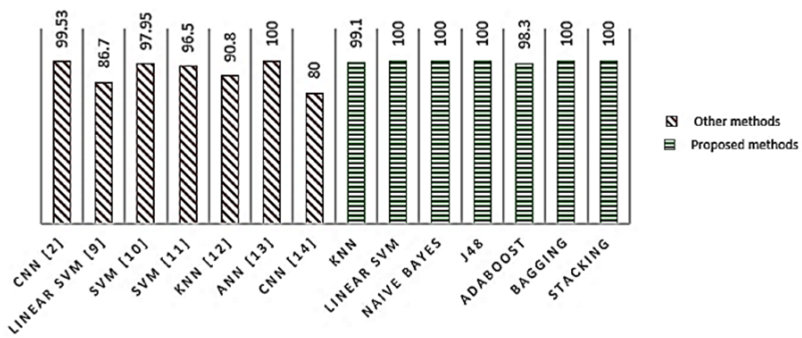

Figure 5: Classification accuracy based on the literature

The comparative result of the SVM method is classified in Figure 14. shows the proposed approach in linear SVM has produced the highest accuracy compared to other methods [1].

Author(s)	Accuracy (%)
B. Li *et al.* [9]	86.7
R. Vaishali and R. Sasikala [10]	97.95
M. Duda *et al.* [11]	96.5
This study	100

Figure 6: Comparison of the classification accuracy using SVM approach

In the proposed "**Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques**" study, Logistic Regression performed the best on the given dataset, with the highest accuracy, precision, recall, and F1 score [3]. Naive Bayes performed well on the dataset, but its accuracy was slightly lower than Logistic Regression. Support Vector Machine and K-Nearest Neighbors performed similarly, with lower accuracy than Logistic Regression and Naive Bayes. Random Forest Classifier performed the worst on the dataset, with the lowest accuracy, precision, recall, and F1 score [3].

	LR	NB	SVM	KNN	RFC
Accuracy	97.15%	94.79%	93.84%	90.52%	81.52%
Confusion matrix	$\begin{bmatrix} 57 & 5 \\ 1 & 148 \end{bmatrix}$	$\begin{bmatrix} 56 & 6 \\ 5 & 144 \end{bmatrix}$	$\begin{bmatrix} 52 & 10 \\ 3 & 146 \end{bmatrix}$	$\begin{bmatrix} 51 & 11 \\ 9 & 140 \end{bmatrix}$	$\begin{bmatrix} 45 & 17 \\ 14 & 135 \end{bmatrix}$
F1 score	0.98	0.96	0.95	0.93	0.88

Figure 7: A comparison of the applied ML models

In the **"Toward Machine Learning-Based Psychological Assessment of Autism Spectrum Disorders in School and Community"** study, the dataset was split into 20% data for testing and 80% data for training [6]. Four machine learning models (SVM, KNN, RF, ANN) were trained and tested on the accumulated dataset. Accuracy and F1-Score were calculated for model performance and comparison [6]. The testing accuracy for SVM, KNN, RF, and ANN was 89%, 78%, 83%, and 89.8%, respectively, with training accuracy near 100% for all classifiers. The achieved F1-Score of SVM, KNN, RF, and ANN in testing data is 86%, 73%, 83%, and 85% subsequently. Figure 15 depicts the comparison of accuracy, recall, precision, and F1-Score among ML classifiers [6]. The evaluation metrics show that SVM and ANN perform significantly better than KNN and RF for ASD classification.

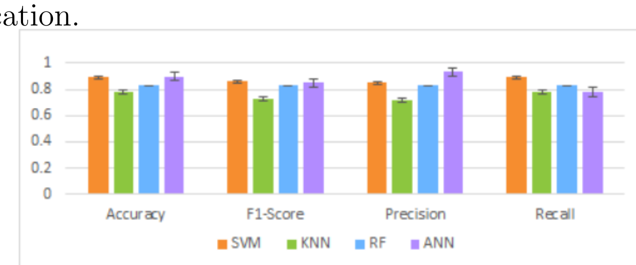

Figure 8: Comparison of performance metrics in testing data among ML classifiers

In "**Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques**" study, the performance of machine learning models was assessed using metrics like specificity, sensitivity, and accuracy from the confusion matrix[2]. Various models were implemented for ASD screening across adults, children, and adolescents, with CNN yielding the highest accuracy across datasets. For adults, CNN achieved a 99.53% accuracy, outperforming other models like K-NN, which had the lowest at 95.75%[2]. In

Classifier	Specificity	Sensitivity	Accuracy
Logistic Regression	0.9575	0.9696	96.69
SVM	0.9574	0.88888	98.11
Naive Bayes	0.9361	96.96	96.22
KNN	0.9148	0.9696	95.75
ANN	0.9787	0.9757	97.64
CNN	1.0	0.9939	99.53

Figure 9: Overall Results for Autistic Spectrum Disorder Screening Data for Adult

the children's dataset, K-NN had the least accuracy at 88.13%, while CNN, SVM, ANN, and LR tied with a 98.30% accuracy rate [2].

Classifier	Specificity	Sensitivity	Accuracy (%)
Logistic Regression	1.0	0.9677	98.30
SVM	1.0	0.9679	98.30
Naive Bayes	0.9642	0.9354	94.91
KNN	0.9642	0.8064	88.13
ANN	0.9642	1.0	98.30
CNN	1.0	0.9678	98.30

Figure 10: Overall Results for Autistic Spectrum Disorder Screening Data for children

For the adolescent dataset, CNN topped with a 96.88% accuracy, whereas K-NN lagged at 80.95% [2].

Classifier	Specificity	Sensitivity	Accuracy (%)
Logistic Regression	1.0	0.6666	85.71
SVM	1.0	0.8888	95.23
Naive Bayes	0.9166	0.8888	90.47
KNN	1.0	0.5555	80.95
ANN	1.0	0.7777	90.47
CNN	1.0	0.9335	96.88

Figure 11: Overall Results for Autistic Spectrum Disorder Screening Data for Adolescent

The experimental results in "**Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder**" assessed F-measure, recall, and precision for three well-known AC algorithms on adult and toddler autism datasets [7]. Different test sizes and label encodings revealed insights into algorithm performance, with varying trends in F1 scores.

- **KNN** outperforms logistic regression and random forest in terms of F-measure, recall, and precision on both the autism adult and toddler datasets [7].

- For **logistic regression**, the F1 score decreases as the test size increases for label 0, but remains constant for label 1 [7].

- For **random forest and KNN**, the F1 score for label 0 decreases until the test size reaches 0.6, and then gradually increases. The F1 score for label 1 remains relatively constant for both algorithms [7].

Accuracy Score is calculated and the accuracy for KNN is 69.2% while accuracy for logistic regression and random forest classifiers are 68.601% and 67.78% respectively [7]. Out of above three models KNN classifier and random forest classifier performs same overall but much better than logistic regression [7].

5 Conclusions

Autism is a growing developmental disorder, and machine learning classification models can help to improve early diagnosis [7]. A significant study for autism is upgrading the performance of the diagnostic forms minimizing the diagnostics time effectively without affecting the validity or sensitivity of the test [1].

The proposed "**Classification of adult autistic spectrum disorder using machine learning approach**" study for adults dataset cross validation was implemented with 3, 5, and 10-folds into the dataset but to evaluate the classification accuracy with other methods in the literature, only 10-fold cross validation was used [1]. According to the results, Bagging, Stacking, linear SVM, naïve bayes, and J48 have achieved a significant accuracy at 100% [1].

In "**Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques**", Logistic regression was the best model for ASD prediction, but more data and deep learning techniques are needed to improve accuracy and robustness[3].

Early ASD diagnosis is crucial for effective intervention, and the **"Toward Machine Learning-Based Psychological Assessment of Autism Spectrum Disorders in School and Community"** study employed a comprehensive approach, including questionnaire-based data collection, correlation analysis, and machine learning classifiers, to address the diverse nature of ASD symptoms. While the machine learning classifiers exhibited good performance, the study acknowledges the need for larger and more diverse datasets, as well as the integration of multi-modal data sources like video, voice, and images to enhance ASD detection in the future [6].

"**Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques**" study collected and analyzed diverse ASD symptoms to create a scenario-based questionnaire, conducted a survey, and employed machine learning classifiers (SVM, KNN, RF, ANN) with good performance [2]. A limitation is the dataset size, prompting the exploration of additional data sources for improved detection [2].

The study "**Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder**" evaluated the performance of three machine learning algorithms on two autism datasets, and found that KNN had the highest accuracy score of 69.2%. This suggests that KNN is a promising algorithm for developing accurate and effective autism detection tools [7].

References

[1] Nurul Amirah Mashudi, Norulhusna Ahmad, and Norliza Mohd Noor. Classification of adult autistic spectrum disorder using machine learning approach. *IAES International Journal of Artificial Intelligence*, 10(3):743, 2021.

[2] Suman Raj and Sarfaraz Masood. Analysis and detection of autism spectrum disorder using machine learning techniques. *Procedia Computer Science*, 167:994–1004, 2020.

[3] Kaushik Vakadkar, Diya Purkayastha, and Deepa Krishnan. Detection of autism spectrum disorder in children using machine learning techniques. *SN Computer Science*, 2:1–9, 2021.

[4] Simon Baron-Cohen, Jane Allen, and Christopher Gillberg. Can autism be detected at 18 months?: The needle, the haystack, and the chat. *The British Journal of Psychiatry*, 161(6):839–843, 1992.

[5] Centers for Disease Control, Prevention, et al. Data & statistics on autism spectrum disorder, 2020.

[6] Sabbir Ahmed, Md Farhad Hossain, Silvia Binte Nur, M Shamim Kaiser, and Mufti Mahmud. Toward machine learning-based psychological assessment of autism spectrum disorders in school and community. In *Proceedings of Trends in Electronics and Health Informatics: TEHI 2021*, pages 139–149. Springer, 2022.

[7] G Devika Varshini and R Chinnaiyan. Optimized machine learning classification approaches for prediction of autism spectrum disorder. *Ann Autism Dev Disord. 2020; 1 (1)*, 1001, 2020.

[8] M Duda, R Ma, N Haber, and DP Wall. Use of machine learning for behavioral distinction of autism and adhd. *Translational psychiatry*, 6(2):e732–e732, 2016.

[9] Osman Altay and Mustafa Ulas. Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and k-nearest neighbor in children. In *2018 6th international symposium on digital forensic and security (ISDFS)*, pages 1–4. IEEE, 2018.

[10] Fadi Thabtah and David Peebles. A new machine learning model based on induction of rules for autism detection. *Health informatics journal*, 26(1):264–286, 2020.

[11] Milan N Parikh, Hailong Li, and Lili He. Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data. *Frontiers in computational neuroscience*, 13:9, 2019.

[12] R Vaishali and R Sasikala. A machine learning based approach to classify autism with optimum behaviour sets. *International Journal of Engineering & Technology*, 7(4):18, 2018.

[13] JA Kosmicki, V Sochat, M Duda, and DP Wall. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational psychiatry*, 5(2):e514–e514, 2015.

[14] Md Delowar Hossain, Muhammad Ashad Kabir, Adnan Anwar, and Md Zahidul Islam. Detecting autism spectrum disorder using machine learning. *arXiv preprint arXiv:2009.14499*, 2020.

[15] FF Thabtah. Uci machine learning repository: Autism screening adult data set, 2017.

[16] Fadi Thabtah. Autism spectrum disorder screening: machine learning adaptation and dsm-5 fulfillment. In *Proceedings of the 1st International Conference on Medical and health Informatics 2017*, pages 1–6, 2017.