RESEARCH ARTICLE OPEN ACCESS

A Novel Method for Detection of Autism Spectrum Disorder Using Machine Learning

[1]G.Divya, [2]Dr.V.Maniraj

[1]Research Scholar, Department Of Computer Science,
A.V.V.M Sri Pushpam College (Autonomous), Poondi,Thanjavur(Dt), Affliated to Bharathidasan University,Thiruchirappalli, Tamilnadu (Mail Id-gdivya19.mca@gmail.com)

[2]Associate Professor, Research Supervisor, Head of the Department,
Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi,Thanjavur(Dt), Affliated To Bharathidasan University,Thiruchirappalli, Tamilnadu,
(Mail Id-manirajv61@gmail.com)

-------------------------------------*************************---------------------------------

Abstract:

Autism Spectrum Disorder (ASD) poses significant challenges to the affected individuals daily lives, making early intervention crucial for mitigating its severity. In this paper, we present a comprehensive framework aimed at evaluating various Machine Learning (ML) techniques for the early detection of ASD) with a focus on real-time prediction through a user-friendly graphical user interface (GUI).These strategies are applied to feature -scaled datasets, which are then subjected to classification using seven different and popular ML algorithms: Ada Boost (AB), Random Forest (RF), k-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), Logistic Regression (LR), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA).We conducted experiments on four standard ASD datasets representing different age groups (Toddlers, Adolescents, Children and Adults). To evaluate the classification outcomes, we employ various statistical measures such as Accuracy, Receiver Operating Curve (ROC) curve, F1 score, Specificity, Sensitivity etc. Through this analysis, the best-performing classification methods for each ASD dataset have been identified. Overall, our purposed framework Demonstrates promising results compared to existing approaches for the early detection of ASD.

Keywords – Autism Spectrum Disorder (ASD), Early detection, Machine learning, Classification, feature Scaling, Statistical Measures.

-------------------------------------*************************---------------------------------

INTRODUCTION

Autism Spectrum Disorder (ASD) emerges early in life, impacting social interactions, communication, and behavior with a wide range of symptoms and severities. While there is no known cure, early intervention and tailored medical care can significantly improve developmental outcomes for those affected by ASD. Diagnosing ASD, however, remains a complex challenge due to its heterogeneous nature and the variability of symptoms. Traditional diagnostic methods rely heavily on behavioral observation and developmental history, which can be time-consuming and subjective to classifiers. Several studies have examined ASD characteristics and diagnostic models, often comparing different algorithms and techniques. Feature selection methods and ensemble learning Recent advancements in technology have led to innovative approaches for ASD detection, with machine learning playing a pivotal role. These approaches include mobile applications that use questionnaires and other data-collection techniques to assess behavioral patterns, providing a convenient platform for early screening. Machine learning algorithms, especially rule-based methods, have shown considerable promise in improving the accuracy and efficiency of ASD classification. Researchers have developed predictive models tailored to various age groups, utilizing a diverse array of algorithms and data encoding strategies to enhance classification accuracy. Additionally, cognitive computing has been explored to identify complex correlations within ASD datasets, further contributing to the development of diagnostic classifiers.

The integration of graphical user interfaces (GUIs) with machine learning-based approaches has opened new avenues for real-time prediction and user-friendly interaction. GUIs in mobile applications enable users to easily input information and receive immediate feedback on the likelihood of ASD, making the diagnostic process more accessible and intuitive. These advancements in technology, along with ongoing research exploring different algorithms and optimization techniques, indicate a promising future for machine learning in ASD diagnosis. The continued evolution of these tools has the potential to lead to earlier and more accurate diagnosis, ultimately resulting in better outcomes for individuals with ASD and their families.

PROPOSED SYSTEM

Four ASD datasets (Toddlers, Adolescents, Children, and Adults) from the publicly available repositories has been collected such as Kaggle and UCI ML [7], [8], [9], [10]. The authors in [13] created the AS Tests Smartphone app for Toddlers, Children, Adolescents, and Adults ASD screening usingQCHAT-10 and AQ-10. The application computes a score of 0 to 10 for every individual, with which the final score is 6 out of 10 which indicates an individual has positives. In addition, ASD data is obtained from the Andressa while open-source databases are developed in order to facilitate research in this area.

MACHINE LEARNING METHOD

In the realm of machine learning for early-stage detection of Autism Spectrum Disorder (ASD), the supervised learning methods are utilized to analyze data and identify patterns indicative of ASD risk. The Algorithms used in this paper would be Ada Boost (AB), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), Logistic Regression (LR), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA).

CLASSIFICATION ALGORITHMS USED

Ada Boost (AB):

Brief overview: Ada Boost is an ensemble learning method that combines multiple weak classifiers to create a strong classifier.

Rationale: Ada Boost is known for its ability to handle imbalanced datasets and has been successful in various classification tasks.

Implementation: Describe how Ada Boost was implemented in your framework, including any specific parameters or configurations used.

Random Forest (RF):

Brief overview: Random Forest is an ensemble learning method that builds multiple decision trees during training and outputs the mode of the classes as the prediction.

Rationale: Random Forest is robust to over fitting and performs well with high-dimensional data, making it suitable for ASD detection.

Implementation: Explain how Random Forest was integrated into your framework and any adjustments made to optimize its performance.

k-Nearest Neighbors (KNN):

Brief overview: k-Nearest Neighbors is a non-parametric method used for classification by finding the majority class among the k nearest neighbors of a data point.

Rationale: KNN is simple yet effective, especially when there is a clear separation between classes, making it a valuable addition to your classification ensemble.

Implementation: Detail how KNN was incorporated into your framework and any considerations for selecting the value of k.

Gaussian Naive Bayes (GNB):

Gaussian Naive Bayes is a probabilistic classifier based on Bayes' theorem with the assumption of independence between features. GNB is computationally efficient and performs well with small datasets, making it suitable for early ASD detection.

Logistic Regression (LR):

Logistic Regression is a linear model used for binary classification by estimating the probability that a given input belongs to a particular class. LR is interpretable and works well with linearly separable data, providing insight into the relationship between features and the target variable.

Support Vector Machine (SVM):

Support Vector Machine is a supervised learning algorithm that finds the hyperplane that best separates classes in a high-dimensional space. SVM is effective in handling complex datasets with non-linear boundaries, making it suitable for ASD detection where classes may not be linearly separable.

Linear Discriminant Analysis (LDA):

Linear Discriminant Analysis is a dimensionality reduction technique that finds the linear combination of features that best separates classes. LDA is effective when the classes are well-separated and normally distributed, providing insight into the discriminative features for ASD detection.

EVALUATION METRICS FOR CLASSIFICATION OUTCOMES

Dataset	No. of Instances	Positive Class	NegativeClass	Male	Female
Toddlers	1054	728	326	735	319
Children	292	76	216	208	84
Adolescents	104	185	107	49	49
Adults	704	508	196	367	337

Table.1 Dataset description

Evaluation Metrics:

This performance of the classification algorithms for early Autism Spectrum Disorder (ASD) detection. These metrics play a crucial role in quantifying the effectiveness of the proposed framework and guiding decision-making processes.

The following metrics were utilized:

Accuracy: Accuracy measures the proportion of correctly classified instances among all instances in the dataset. It provides an overall assessment of the model's predictive capability but may be misleading in the presence of class imbalance.

$$\%\text{Accuracy} = \frac{\text{No. of correct predictions}}{\text{Total No. of predictions}} * 100$$

Receiver Operating Characteristic (ROC) curve: The ROC curve is a graphical representation of the true positive rate (sensitivity) against the false positive rate (1 - specificity) for different threshold values. It illustrates the trade-off between sensitivity and specificity and is particularly useful for assessing binary classification models.

F1 Score: The F1 score is the harmonic mean of precision and recall, providing a balanced measure of a classifier's performance. It is especially valuable in the presence of class imbalance, as it considers both false positives and false negatives

$$\text{F1 score} = \frac{2(\text{Precision X Recall})}{2a(\text{Precision} + \text{Recall})}$$

Specificity: Specificity measures the proportion of true negative instances correctly identified by the classifier. It is particularly relevant in medical diagnosis scenarios, where minimizing false positives is critical.

$$\text{Specificity} = \frac{TN}{TN+FP}$$

Sensitivity: Sensitivity, also known as recall or true positive rate, measures the proportion of true positive instances correctly identified by the classifier. It is essential for capturing the ability of the model to detect positive cases accurately.

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

RESULTS

The following are the results of various Machine learning Algorithms for early detection of Autism Spectrum disorder on four prominent datasets.

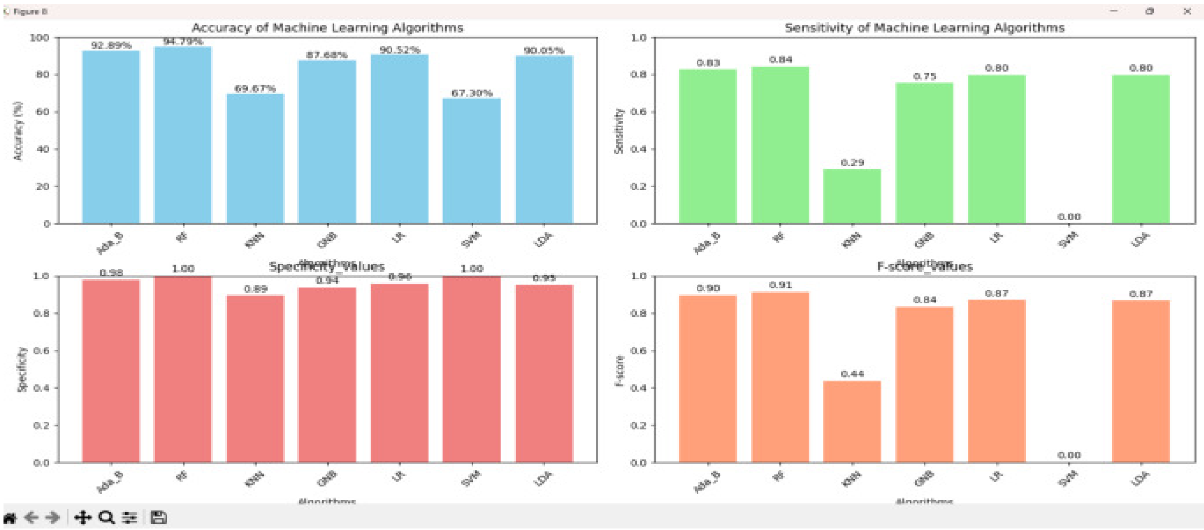

Fig 5.1: Comparative Analysis of Accuracy, Sensitivity, Specificity and F-score for different ML techniques on children Dataset

S. No.	MODEL	%ACCURACY	SENSITIVITY	SPECIFICITY	F-SCORE	CONFUSIONMATRIX
1.	Adaboost	96.21	0.90	0.99	0.94	[[606] 8 137]]
2.	Random forest	95.73	0.90	0.98	0.94	[[588] [3 142]]
3.	KNN	67.77	0.25	0.85	0.39	[[1254] [23 122]]
4.	Gaussian naïvebayes	91.47	0.83	0.95	0.89	[[5214] [9136]]
5	Logistic regression	91.94	0.85	0.95	0.90	[[5013] [6129]]
6	Support vectormachine	71.56	0.00	1.00	0.00	[[3416] [716]]
7	Linear discriminantanalysis	90.52	0.85	0.93	0.89	[[5610] [8137]]

Performance metrics of different ML models

Our study proposes a Autism Spectrum Disorder (ASD) through Machine Learning (ML) techniques, coupled with a user-friendly graphical user interface (GUI), exhibits promising results. By employing feature-scaled datasets and applying seven different ML algorithms—Ada Boost (AB), Random Forest (RF), k-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), Logistic Regression (LR), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA)—we conducted experiments on four standard ASD datasets representing various age groups. Through rigorous evaluation using statistical measures such as Accuracy, Receiver Operating Curve (ROC) curve, F1 score, Specificity, and Sensitivity, we identified the best-performing classification methods for each dataset. The outcomes indicate the effectiveness of our framework in early ASD detection, surpassing existing approaches. This underscores the significance of leveraging ML techniques for timely intervention and management of ASD, potentially improving the quality of life for affected individuals

FUTURE SCOPE

The future scope of autism spectrum disorder (ASD) research and treatment encompasses a variety of promising developments aimed at improving diagnosis, understanding, and support for individuals with ASD and their families. In the coming years, one key area of focus is early detection and diagnosis. Researchers are working on identifying reliable biomarkers and developing advanced neuroimaging techniques, as well as leveraging artificial intelligence (AI) and machine learning to detect early signs of autism in infants and young children. Early diagnosis is crucial because it enables earlier intervention, which can significantly impact developmental outcomes. Another area of ongoing research is genetic and genomic studies, where scientists are exploring the genetic factors and variations associated with ASD.

REFERENCE

[1] M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni, ''Efficient machine learning models for early stage detection of autism spectrum disorder,'' Algorithms, vol. 15, no. 5, p. 166, May 2022.
[2] D. Petrucci, A. Teofani, M. Milanesi, B. Fosso, L. Putignani, F. Messina, G. Pesole, A. Desideri, and G. Chillemi, ''Machine learning data analysis highlights the role of parasutterella and alloprevotella in autism spectrum disorders,'' Biomedicines, vol. 10, no. 8, p. 2028, Aug. 2022.
[3] R. Sreedasyam, A. Rao, N. Sachidanandan, N. Sampath, and S. K. Vasudevan, ''Aarya—A kinesthetic companion for children with autism spectrum disorder,'' J. Intell. Fuzzy Syst., vol. 32, no. 4, pp. 2971–2976, Mar. 2017.
[4] J. Amudha and H. Nandakumar, ''A fuzzy based eye gaze point estimation approach to study the task behavior in autism spectrum disorder,'' J. Intell. Fuzzy Syst., vol. 35, no. 2, pp. 1459–1469, Aug. 2018. \
[5] H. Chahkandi Nejad, O. Khayat, and J. Razjouyan, ''Software development of an intelligent spirography test system for neurological disorder detection and quantification,'' J. Intell. Fuzzy Syst., vol. 28, no. 5, pp. 2149–2157, Jun. 2015.
[7] Kaggle. (2022). Autism Spectrum Disorder Detection Dataset for Toddlers.
[Online]. Available: https://www.kaggle.com/fabdelja/autism-screeningfor-toddlers
[8] UCI. (2022). UCI Machine Learning Repository: Autistic Spectrum Disorder Screening Data for Adolescent Data Set. [Online]. Available:
https://shorturl.at/fhxCZ
[9] UCI. (2022). UCI Machine Learning Repository: Autism Screening Adult Data Set. [Online]. Available: https://archive.ics.uci.edu/
ml/datasets/Autism+Screening+Adult
[10] UCI. (2022). UCI Machine Learning Repository: Autistic Spectrum Disorder Screening Data for Children Data Set. [Online]. Available:https://shorturl.at/fiwLU