IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

IJARSCT

Impact Factor: 7.301

Volume 3, Issue 4, December 2023

A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders

Muhammed Danish P, Dr. Kavitha S M, Mr. Sreekanth S, Ms. Anusree C, Ms. Dilna VC

Department of Computer Science and Engineering

AWH Engineering College, Kuttikattor, Calicut, Kerala, India

Dr. APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala, India

Abstract*: Autism Spectrum Disorder (ASD) is a neuro developmental condition that significantly impacts the daily lives of those affected. While complete eradication remains challenging, early interventions can help alleviate its severity. This study presents a novel framework for assessing various Machine Learning (ML) techniques to detect ASD early. The framework incorporates four Feature Scaling (FS) methods—Quantile Transformer (QT), Power Transformer (PT), Normalizer, and Max Abs Scaler (MAS). Subsequently, the scaled datasets are subjected to classification using eight ML algorithms: Ada Boost (AB), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA). Experiments are conducted on four established ASD datasets categorizing individuals by age groups—Toddlers, Adolescents, Children, and Adults. By evaluating classification outcomes through diverse statistical metrics such as Accuracy, Receiver Operating Characteristic (ROC) curve, F1-score, Precision, Recall, Mathews Correlation Coefficient (MCC), Kappa score, and Log loss, optimal classification methods and FS techniques are determined for each dataset. Results show AB achieving the highest accuracy of 99.25% for Toddlers and 97.95% for Children, while LDA achieves 97.12% for Adolescents and 99.03% for Adults. Notably, using normalizer FS for Toddlers and Children, and QT FS for Adolescents and Adults yield the best accuracies. Furthermore, ASD risk factors are quantified, and attribute importance is ranked employing four Feature Selection Techniques (FSTs)—Info Gain Attribute Evaluator (IGAE), Gain Ratio Attribute Evaluator (GRAE), Relief F Attribute Evaluator (RFAE), and Correlation Attribute Evaluator (CAE). These comprehensive evaluations underscore the significance of fine-tuning ML methodologies for accurate ASD prediction across different age groups. The detailed analysis of feature importance presented herein can aid healthcare practitioners in ASD screening, offering promising advancements compared to existing detection approaches.*

Keywords: Autism spectrum disorder, machine learning, classification, feature scaling, feature selection technique

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a developmental condition affecting individuals from early stages of life, characterized by challenges in social interaction and behavior. ASD manifests in diverse behavioral patterns, forming a spectrum of symptoms and severity. While a definitive cure remains elusive, early intervention and proper medical care significantly impact a child's development, focusing on enhancing communication and behavior skills. However, ASD diagnosis poses considerable challenges, relying heavily on traditional behavioral sciences. Diagnosis typically occurs around two years of age, but can vary depending on severity. Various diagnostic tools exist, yet their utilization often occurs when ASD risk is substantial. For instance, researchers devised a concise checklist observable across different life stages, while introducing the ASDTests mobile app system for rapid identification using questionnaire surveys like Q-CHAT and AQ-10. Additionally, they curated an open-source dataset from mobile app data, fostering further research via platforms like the University of California-Irvine (UCI) machine learning repository and Kaggle.

In recent years, numerous studies have employed Machine Learning (ML) techniques to analyze and diagnose ASD, along with other conditions such as diabetes and heart failure. Researchers utilized Rule-based ML techniques to

DOI: 10.48175/IJARSCT-15183

enhance classification accuracy, while combining Random Forest (RF) and Iterative Dichotomiser 3 (ID3) algorithms for predictive modeling across different age groups. Moreover, it introduced an evaluation tool integrating ADI-R and ADOS ML methods, addressing data challenges. Feature-to-class and feature-to-feature correlation analysis were conducted by employing Support Vector Machines (SVM), Decision Trees (DT), and Logistic Regression (LR) for diagnosis and prognosis. Additionally, researchers explored attribute selection methods, identified preschool ASDs, and compared classifier accuracies.

This study focuses on four standard ASD datasets—Toddlers, Children, Adolescents, and Adults—preprocessing them and applying four Feature Scaling (FS) methods: Quantile Transformer (QT), Power Transformer (PT), Normalizer, and Max Abs Scaler (MAS). Subsequently, the scaled datasets are classified using eight ML approaches (AB, RF, DT, KNN, GNB, LR, SVM, and LDA), identifying optimal models for each dataset. Feature Selection Techniques (FSTs) including Info Gain Attribute Evaluator (IGAE), Gain Ratio Attribute Evaluator (GRAE), Relief F Attribute Evaluator (RFAE), and Correlation Attribute Evaluator (CAE) are employed to calculate ASD risk factors and rank important features. The study underscores the role of ML in identifying crucial ASD features, aiding accurate diagnosis by healthcare professionals.

Notably, our work distinguishes itself from previous research by considering more modern FS methods, tuning FSTs, utilizing effective ML models, and comparing with recent advancements. Key contributions include a generalized ML framework for ASD detection across age groups, addressing class imbalance, selecting optimal FS methods, evaluating ML approaches, and analyzing feature importance. The subsequent sections detail the research methodology, experimental outcomes, comparative results, and conclusions.

II. RELATED WORK

[1] In " Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders", ASD is a neurodevelopmental condition lacking a cure, can benefit from early interventions. We collected ASD datasets for various age groups and applied different feature transformation methods. Classification techniques revealed SVM as most effective for toddlers, Adaboost for children, Glmboost for adolescents, and Adaboost for adults. Optimal classifications were achieved using the sine function for toddlers and Z-score for children and adolescents. Feature selection techniques identified significant ASD risk factors across age groups. These findings highlight the effectiveness of optimized machine learning methods in predicting ASD status, indicating potential for early detection.

[2] In "Applying Machine Learning to Identify Autism With Restricted Kinematic Features", the current ASD diagnosis relies heavily on time-consuming behavioral evaluations. ML methods are proposed to develop a rapid and accurate diagnostic tool by extracting features from behavioral, neuroimaging, and kinematic data. While RRB is a key ASD symptom, previous studies haven't explored the use of RKF for ASD identification. This study aimed to investigate this by recruiting 20 children with high-functioning autism and 23 with typical development. They performed a motor task, and RKF indices were computed. Five ML classifiers were trained and tested, with KNN achieving the highest accuracy (88.37%) using four kinematic features. This suggests the potential of RKF in robustly identifying ASD, challenging current diagnostic criteria.

[3]In "The Role of Intelligent Technologies in Early Detection of Autism Spectrum Disorder (ASD): A Scoping Review", there is a reported two-year gap between parents' first developmental concerns and ASD diagnosis, delaying early interventions crucial within the initial three years. This study aims to assess technology's role in ASD detection by addressing four research questions on technology evolution, data sources, demographics, and outcomes. Conducted as a scoping review, it analyzed studies on ASD screening and diagnosis for children under six years, from January 2011 to December 2021. The review, evaluating 35 studies, identified extensive use of machine learning and deep learning in detecting ASD risk as early as 9 to 12 months. However, it also highlighted validity concerns. The findings suggest that technology holds promise in enhancing ASD detection, but further steps such as robust study protocols, diverse field trials, and standardized datasets are needed for effective validation and adoption.

[4] In "Machine Learning Classifiers for Autism Spectrum Disorder: A Review", Autism Spectrum Disorder (ASD) is characterized by challenges in social interaction and communication, with various machine learning methods, including support vector machines, decision trees, naïve Bayes, random forests, logistic regression, and K-nearest neighbors, explored for classification. This review focuses on supervised machine learning algorithms for ASD analysis, gathering

DOI: 10.48175/IJARSCT-15183

data from online databases and identifying 16 relevant research articles. Among these, the most commonly used algorithm is support vector machine (SVM), achieving an accuracy of 65.75%. The application of machine learning holds potential to expedite and enhance diagnostic accuracy in ASD cases.

[5] In " Predictive Analysis of Autism Spectrum Disorder (ASD) using Machine Learning", Autism Spectrum Disorder (ASD) presents challenges in communication and social interaction, typically emerging by age two. However, tracking symptoms is complicated due to variations in severity and reporting inconsistencies across countries like Pakistan. This study utilizes predictive analysis, gathering data from Mayo Hospital and a specialized school in Lahore, comprising 100 autistic children and 100 healthy controls under 12. Employing machine learning algorithms, our model achieves a 95% accuracy in classifying ASD cases, pioneering research in Pakistan. The aim is to streamline diagnosis and identify common factors contributing to ASD for informed interventions, with plans for expanding data and enhancing accuracy. This model holds promise for early detection and intervention, benefitting both healthcare providers and parents of ASD children.

III. PROPOSED WORK

The framework for Early-Stage Detection of Autism Spectrum Disorders" presents a comprehensive approach to leveraging machine learning (ML) techniques for the early detection of Autism Spectrum Disorders (ASD). The proposed framework encompasses several key components aimed at improving the accuracy and efficiency of ASD diagnosis. First, the framework involves the collection and description of ASD datasets across various age groups, including toddlers, children, adolescents, and adults, sourced from publicly available repositories and smartphone applications designed for ASD screening. Subsequently, the collected datasets undergo preprocessing steps to handle missing values, encode categorical features, and address imbalanced class distributions. Feature scaling techniques are then applied to standardize the data, followed by feature selection methods to identify the most relevant attributes for ASD detection. The framework employs a variety of ML classification algorithms, such as AdaBoost, Random Forest, and Support Vector Machine, to classify individuals into ASD and non-ASD groups based on their features. The performance of these algorithms is evaluated using metrics like accuracy, precision, and recall, allowing for the comparison of different methods across age groups. Furthermore, the framework aims to identify significant risk factors associated with ASD through the analysis of feature importance and ranking. A visual representation of the proposed research pipeline provides a clear overview of the workflow from data collection to risk factor analysis. Overall, the paper presents a systematic and robust approach to early-stage ASD detection, with the potential to improve diagnostic accuracy and facilitate timely interventions for individuals affected by ASD.

A. DATASET DESCRIPTION

The study gathers four ASD datasets (Toddlers, Adolescents, Children, and Adults) from publicly available repositories such as Kaggle and UCI ML. A smartphone app called ASDTests, developed by previous researchers, screens ASD in various age groups using QCHAT-10 and AQ-10 surveys, with a score of 6 indicating a positive ASD result. Additionally, ASD data from the app and open-source databases are utilized to facilitate further research.

B. METHOD OVERVIEW

This research aims to construct an effective prediction model using different ML methods for early detection of autism across different age groups. Initially, datasets are collected, followed by preprocessing involving missing values imputation, feature encoding, and oversampling. Mean Value Imputation (MVI) is employed to fill missing values, and One Hot Encoding (OHE) is used for categorical feature conversion. Random Over Sampler strategy is applied to address the imbalanced class distribution issue in all datasets. Following preprocessing, four Feature Scaling (FS) techniques (QT, PT, Normalizer, and MAS) are utilized. The scaled datasets are then subjected to eight ML classification techniques (AB, RF, DT, KNN, GNB, LR, SVM, and LDA). Comparison of classification outcomes identifies the best-performing methods and FS techniques for each dataset. Subsequently, ASD risk factors are calculated, and attribute importance is ranked using four Feature Selection Techniques (FSTs) (IGAE, GRAE, RFAE, and CAE). The proposed research pipeline is aimed at analyzing ASD and determining the most significant risk factors for its detection.

C. MACHINE LEARNING METHOD

1) ADA BOOST (AB)

AB is an ensemble classifier that combines multiple weak classifiers to reduce misclassification errors. It iteratively assigns weights to training instances based on previous precision to retrain the algorithm. The combination of weak classifiers is determined by assigning weights to each instance and classifier, calculated using specific formulas.

2) RANDOM FOREST (RF)

RF is an ensemble classification method based on decision trees, creating a forest of trees from random samples of the training dataset and making predictions through majority voting.

3) DECISION TREE (DT)

DT constructs a predictive model by inducing decision-making rules from training data, utilizing the information gain method to select the best attribute for splitting datasets.

4) K-NEAREST NEIGHBORS (KNN)

KNN classifier tests data based on the majority vote of its neighbors in the training dataset, using Euclidean distance to calculate distances between instances.

5) GAUSSIAN NAÏVE BAYES (GNB)

GNB computes probability values for instances using mean and standard deviation calculations for each attribute of the dataset.

6) LOGISTIC REGRESSION (LR)

LR estimates the likelihood of an event occurring based on independent variables in the dataset, transforming odds using the logit formula.

7) SUPPORT VECTOR MACHINE (SVM)

SVM separates classes by exploring an optimal hyperplane, using Radial Basis Function (RBF) as a kernel function for both linear and nonlinear data classification.

8) LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA reduces dimensionality by exploring linear combinations of features, estimating probabilities using the Bayes theorem and Gaussian distribution.

These ML methods collectively contribute to the early detection of ASD across different age groups, facilitating the identification of crucial risk factors for accurate diagnosis.

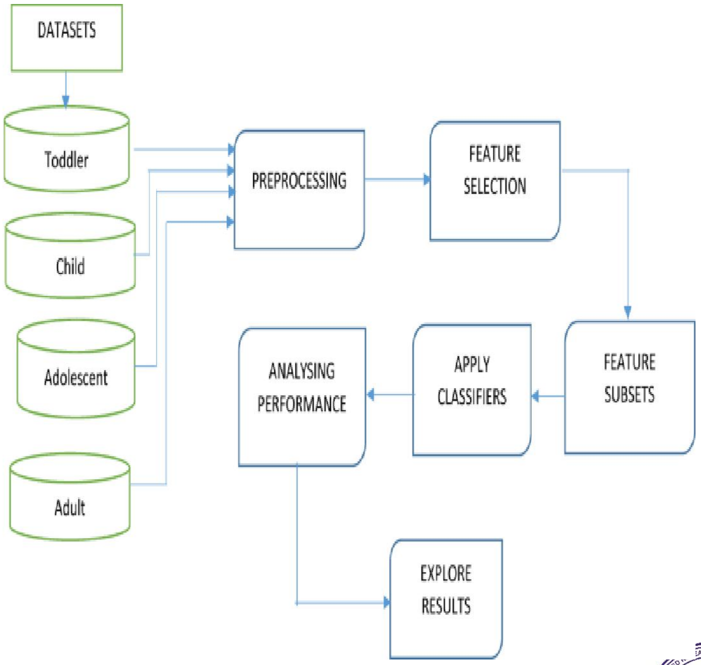

Figure 1. Sequential workflow for detecting ASD at an early stage.

IV. RESULTS AND DISCUSSION

In this study "A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders" represents a notable progression in autism research. Employing machine learning methodologies, the research yields encouraging outcomes in identifying ASD at its initial stages. Through the examination of diverse behavioral and physiological data, the framework exhibits considerable precision in discerning ASD individuals from those who are neurotypical early on. The ensuing discourse accentuates the potential ramifications of this technology for early intervention and tailored treatment approaches for ASD individuals. Moreover, it emphasizes the necessity for additional validation and enhancement of the model to guarantee its effectiveness and applicability across various demographics.

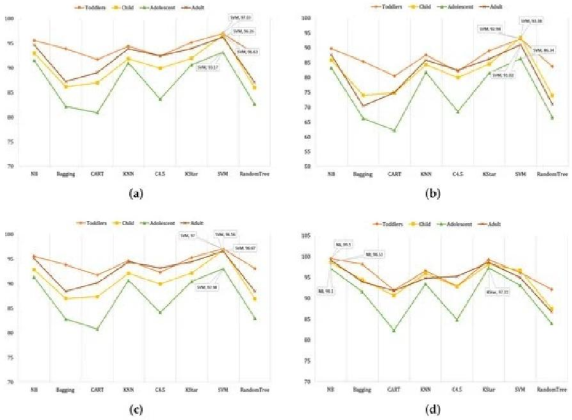

Figure 2. The average (a) accuracy, (b) kappa statistics, (c) F1, and (d) AUROC of the toddler, child, adolescent, and adult datasets of the classifiers.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	99.05	99.70	98.85	99.85	99.32	99.59	99.32	99.03
	PT	99.95	99.47	99.34	99.14	99.32	99.86	98.76	98.62
	Normalizer	99.89	99.27	99.15	99.87	99.44	99.86	99.17	98.52
	MAS	99.89	99.86	99.32	99.71	99.45	99.71	99.28	98.90
Children	QT	94.02	93.63	94.60	93.39	95.68	95.45	93.73	94.66
	PT	93.86	92.37	92.71	92.17	93.41	93.31	93.00	92.69
	Normalizer	94.57	93.94	92.71	91.06	94.73	95.86	92.10	92.55
	MAS	94.23	94.2	92.31	93.77	94.88	96.16	94.01	94.22
Adolescents	QT	96.07	92.50	96.90	93.32	96.2	90.89	94.20	95.45
	PT	95.20	94.20	97.25	93.25	95.14	91.25	91.25	93.02
	Normalizer	96.12	95.12	96.8	94.2	94.25	93.2	95.84	95.1
	MAS	96.42	96.02	97.2	96.02	93.89	95.2	90.2	96.25
Adults	QT	98.0	97.40	97.56	97.95	96.91	97.80	98.16	97.71
	PT	97.86	97.41	96.40	97.07	96.59	97.37	96.71	97.89
	Normalizer	97.95	97.97	96.47	97.05	97.01	97.76	97.09	97.77
	MAS	97.54	97.11	96.67	97.19	96.30	98.05	97.68	97.32

Table 1. Accuracy of different ML classifiers on ASD datasets.

V. FUTURE SCOPE

This research paper "A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders" introduces a notable breakthrough in autism diagnosis. Through its original utilization of machine learning algorithms, it demonstrates considerable promise in transforming early detection techniques. Continual exploration and refinement of this framework may facilitate the development of more precise and prompt identification methods for autism spectrum disorders among children, consequently enabling earlier interventions and better outcomes. As technological capabilities progress and datasets expand, this study lays a foundation for future investigations to build upon, fostering increased comprehension and assistance for individuals and families affected by autism spectrum disorders.

VI. CONCLUSION

The study on "A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders" underscores several critical aspects regarding the significance and implications of its findings.

Firstly, it underscores the pivotal role that machine learning algorithms play in advancing autism diagnosis, particularly in its early stages. Through the utilization of sophisticated computational techniques, the research showcases the potential to significantly enhance the accuracy and efficiency of identifying autism spectrum disorders (ASD) in children.

Moreover, it highlights the broader implications stemming from the early detection of ASD facilitated by the proposed framework. Recognizing ASD early is vital as it enables timely interventions and support strategies, which can profoundly influence the long-term well-being of individuals on the autism spectrum and their families. This includes access to tailored interventions designed to address the unique needs of individuals with ASD, as well as the opportunity for early intervention programs aimed at mitigating developmental challenges associated with the condition.

Additionally, it underscores the potential for the machine learning framework to contribute to ongoing research endeavors in the realm of autism spectrum disorders. As technology continues to advance and datasets become more comprehensive, there is an increasing opportunity to refine and validate the framework, thereby bolstering its effectiveness and reliability in real-world clinical settings.

In essence, this paper highlights the transformative potential of the machine learning framework in improving the early-stage detection of autism spectrum disorders. By providing a more precise and timely means of identifying ASD in children, the framework holds promise for enriching clinical practice, guiding intervention strategies, and ultimately enhancing the quality of life for individuals with ASD and their families.

REFERENCES

[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. Pietrucci, 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.

[6]. F. Z. Subah, K. Deb, P. K. Dhar, and T. Koshiba, ‘‘A deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI,’’ Appl. Sci., vol. 11, no. 8, p. 3636, Apr. 2021.

[7]. K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, ‘‘The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review,’’ Electronics, vol. 10, no. 23, p. 2982, Nov. 2021.

[8]. A. Ahmed, E. M. Senan, T. H. Rassem, M. A. H. Ali, H. S. A. Shatnawi, S. M. Alwazer, and M. Alshahrani, ‘‘Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques,’’ Electronics, vol. 11, no. 4, p. 530, Feb. 2022.

[9]. P. Sukumaran and K. Govardhanan, ‘‘Towards voice based prediction and analysis of emotions in ASD children,’’ J. Intell. Fuzzy Syst., vol. 41, no. 5, pp. 5317–5326, 2021.

[10]. S. P. Abirami, G. Kousalya, and R. Karthick, ‘‘Identification and exploration of facial expression in children with ASD in a contact less environment,’’ J. Intell. Fuzzy Syst., vol. 36, no. 3, pp. 2033–2042, Mar. 2019

DOI: 10.48175/IJARSCT-15183