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
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

Review Paper on An Early-Stage Autism Spectrum Detection System

Prof. D. C. Pardeshi[1], Parth S. Mishra[2], Shripad Kulkarni[3]
Professor, Department of Artificial Intelligence and Machine Learning[1]
Students, Department of Artificial Intelligence and Machine Learning[2,3]
All India Shri Shivaji Memorial Society Polytechnic Pune, Maharashtra, India

Abstract*: The Early-Stage Autism Detection System presents a breakthrough approach to identifying Autism Spectrum Disorder (ASD) in its initial stages, particularly focusing on early childhood diagnosis. Leveraging machine learning (ML) techniques such as Random Forest and Support Vector Machines, the system meticulously analyses behavioural patterns and social interactions to pinpoint potential indicators of ASD, even in toddlers. It adeptly tackles challenges like imbalanced class distributions by employing random oversampling and adopts feature scaling and selection methods to heighten prediction accuracy. Through extensive experimentation on diverse ASD datasets, the system discerns crucial features pivotal for precise diagnosis. Its implementation promises timely intervention and improved outcomes by enabling the early detection and support of individuals with ASD from the outset of development.*

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

I. INTRODUCTION

Autism Spectrum Disorder (ASD) poses significant challenges in early identification and intervention due to its varied symptoms and impacts on social interactions and behaviors. Despite lacking a sustainable solution, early intervention and proper medical care can significantly improve a child's development, focusing on enhancing communication skills and behavioural patterns. Traditional methods of ASD diagnosis rely on behavioural science and are often complex and difficult. Typically, ASD is diagnosed around the age of two, although it can be identified later based on severity. Various treatment strategies exist, but they aren't universally applied until ASD development becomes severe.In recent years, researchers have turned to machine learning (ML) approaches to aid in ASD diagnosis.

These studies have utilized a range of ML algorithms, including rule-based ML techniques, Random Forest (RF), Support Vector Machines (SVM), Decision Trees (DT), and others. By analyzing ASD attributes, researchers have developed predictive models for different age groups, from toddlers to adults. The integration of cognitive computing has enabled the identification of significant features for ASD diagnosis.

Furthermore, studies have examined the predictive performance of ML models such as Deep Neural Networks (DNN), ensemble ML approaches, and bio-inspired algorithms. These efforts have resulted in the development of smartphone applications and programming interfaces for ASD diagnosis across all age groups.

To enhance prediction accuracy, researchers have addressed imbalanced class distributions in ASD datasets and investigated various feature scaling methods. Additionally, feature selection techniques have been employed to identify the most important risk factors for ASD prediction. One notable study has compared different ML models and feature selection methods to identify the most effective approaches for ASD detection. By conducting extensive experiments on standard ASD datasets, researchers have aimed to improve the accuracy and reliability of ASD diagnosis.

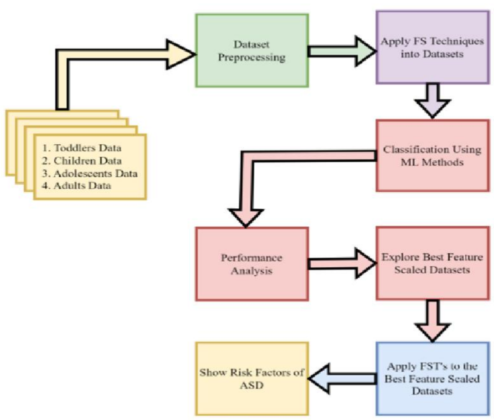
In summary, the application of ML techniques in ASD diagnosis shows promise for early-stage detection and intervention. By leveraging diverse datasets and advanced algorithms, researchers aim to enhance our understanding of ASD and provide better support for affected individuals across the lifespan. Ongoing research in this area continues to refine ML models and improve their effectiveness in ASD diagnosis and management

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
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

II. MATERIALS & METHODS

DATASET DESCRIPTION: We collect the four ASD datasets (Toddlers, Adolescents, Children, and Adults) from the publicly available repositories: Kaggle and UCI ML [36], [37], [38], [39]. The authors in [13] created the ASD Tests smartphone app for Toddlers, Children, Adolescents, and Adults ASD screening using QCHAT-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 positive ASD.

METHOD OVERVIEW: This research aims to create an effective prediction model using different types of ML methods to detect autism in people of different ages. First of all, the datasets are collected, and then the preprocessing is accomplished via the missing values imputation, feature encoding, and oversampling. The Mean Value Imputation (MVI) method is used to impute the missing values of the dataset. Then, the categorical feature values are converted to their equivalent numerical values using the One Hot Encoding (OHE) technique. Table 1 shows that all four datasets used in this work have an imbalanced class distribution problem. As such, a Random Over Sampler strategy is used to alleviate this issue. After completing the initial preprocessing, the datasets feature values are scaled using four different FS techniques i.e., QT, PT, Normalizer, and MAS (see their detailed operations in Table 3). The feature-scaled datasets are then classified using eight different ML classification techniques i.e., AB, RF, DT, KNN, GNB, LR, SVM, and LDA. Comparing the classification outcomes of the classifiers on different feature-scaled ASD datasets, the best-performing classification methods, and the best FS techniques for each ASD dataset are identified. After those analyses, the ASD risk factors are calculated, and the most important attributes are ranked according to their importance values using four different FSTs i.e., IGAE, GRAE, RFAE, and CAE. To this end, Fig. 1 represents the proposed research pipeline to analyse the ASD datasets and calculate the risk factors that are most responsible for ASD detection.



MACHINE LEARNING METHODS:

1) *ADA BOOST (AB)*: AB is a tree-based ensemble classifier that incorporates many weak classifiers to reduce misclassification errors.

2) *RANDOM FOREST (RF)*: RF is a decision tree-based ensemble classification method and follows the split and conquer technique in the input dataset to create multiple decision-making trees (known as the forest) [42]. It works in two phases. At first, it creates a forest by combining the 'N' number of decision trees and in the second phase, it makes predictions for each tree generated in the first phase. The working process of the RF algorithm is illustrated below:

Select random samples from the training dataset.
Construct decision trees for each training sample.
Select the value of N to define the number of decision trees.
Repeat Steps 1 and 2.
For each test sample, find the predictions of each decision tree, and assign the test sample a class value based on majority voting.

3) *DECISION TREE(DT)*: DT follows a top-down approach to build a predictive model for class values using training data-inducing decision-making rules[43].This research utilized the information gain method to select the best attribute.

Assuming Pi, the probability such that $xi \in D$, exists to a class Ci, and is predicted by $|Ci\ D|/|D|$. To classify instances in the dataset D, the required information Is needed.

4) *GAUSSIAN NAIVE BAYES (GNB)*: GNB algorithm follows a normal distribution and is used for classification when all the data values of a dataset are numeric [43]. To compute the probability values of any instance with respect to the class value mean and standard deviation are calculated for each attribute of the dataset. Consequently, for testing, when any instance comes, it utilizes the mean and standard deviation values to calculate the probability of the test instance.

5) *K-NEAREST NEIGHBORS (KNN)*: KNN classifies the test data by utilizing the training data directly by calculating the K value, indicating the number of KNN [43]. For each instance, it computes the distance between all the training instances and sorts the distance. Furthermore, a majority voting technique is employed to assign the final class label to the test data. This research applies Euclidean distance to calculate the distances among instances

III. RESULTS AND REVIEWS

We analysed four different ASD datasets to build prediction models for different stages of people. In order to do this, we applied various FS methods to those ASD datasets and classified them utilizing eight different simple but effective ML classifiers and also determined how the FS methods affect the classification performance. Furthermore, we also employed four different FSTs to compute the importance of the features which are more responsible for ASD prediction. Inspecting the experimental findings, the best performing classifiers model predicted ASD with AB (99.25%), AB (97.95%), LDA (97.12%), LDA (99.03%) accuracy; AB, LR (99.99%), GNB

IV. CONCLUSION

In this study, we developed a machine-learning framework to detect Autism Spectrum Disorder (ASD) across different age groups (Toddlers, Children, Adolescents, and Adults) using predictive models based on various ML techniques. Following initial data processing, ASD datasets underwent scaling using four feature scaling techniques (QT, PT, Normalizer, MAS) and classification using eight ML classifiers (AB, RF, DT, KNN, GNB, LR, SVM, LDA). We evaluated classification performance using accuracy, ROC, F1-Score, precision, recall, MCC, kappa score, and Log loss metrics to determine optimal FS and classification methods. Our ML-based prediction models offer accurate ASD identification, aiding physicians in diagnosis. We also analyzed feature importance using IGAE, GRAE, RFAE, and CAE, identifying key ASD prediction features. While data limitations hindered a fully generalized model, future work will gather more diverse ASD data to enhancedetection across all ages and neuro-developmental disorders.

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. 29712976, Mar. 2017.

[4]. J. Amudha and H. Nandakumar, A fuzzy based eye gaze point estimation approach to study the task behavior in autism spec trum disorder, J. Intell. Fuzzy Syst., vol. 35, no. 2, pp. 14591469, 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. 21492157, 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.

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
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

[8]. I. 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. 53175326, 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. 20332042, Mar. 2019.
[11]. M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, Detecting autism spectrum disorder using machine learning techniques, Health Inf. Sci. Syst., vol. 9, no. 1, pp. 1 13, Dec. 2021.
[12]. C. Allison, B. Auyeung, and S. Baron-Cohen, Toward brief red flags for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls, J. Amer. Acad. Child Adolescent Psychiatry, vol. 51, no. 2, pp. 202212, 2012.
[13]. F. Thabtah, F. Kamalov, and K. Rajab, A new computational intelligence approach to detect autistic features for autism screening, Int. J. Med. Inform., vol. 117, pp. 112124, Sep. 2018.
[14]. E. Dritsas and M. Trigka, Stroke risk prediction with machine learning techniques, Sensors, vol. 22, no. 13, p. 4670, Jun. 2022