



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

A Machine Learning Framework for Advance Phase Identification of Autism Spectrum Disorders

Mrs. J. Ramya, Preethi S, Selvapriya P, Sneha V

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT: Autism Spectrum Disorder (ASD) is a neurological impairment condition that severely impairs cognitive, linguistic, object recognition, interpersonal, and communication skills. Its main cause is genetic, and early treatment and identification can reduce the patient's expensive medical costs and lengthy examinations. We developed a machine learning (ML) architecture that is capable of effectively analysing autistic children's datasets and accurately classifying and identifying ASD traits. We considered the ASD screening dataset of toddlers in this study. We utilised the SMOTE method to balance the dataset, followed by feature transformation and selection methods. Then, we utilised several classification techniques in conjunction with a hyperparameter optimisation approach. The AdaBoost method yielded the best results among the classifiers. We employed ML and statistical approaches to identify the most crucial characteristics for the rapid recognition of ASD patients. We believe our proposed framework could be useful for early diagnosis and helpful for clinicians.

KEYWORDS: Autism Spectrum Disorder; Machine Learning; Feature Transformation; Feature Selection; Hyper-Parameter Optimization

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a group of neurological diseases that hinder the normal progression of the brain [1]. ASD causes social difficulties, sensory issues, repetitive behaviour, and intellectual incapacity. ASD patients also have psychiatric or neurological disorders such as hyperactivity, attention deficit, anxiety, sadness, and epilepsy [2]. ASD has been connected to genetics and physiological factors. However, the disorder is often assessed utilizing non-genetic behavioural criteria. Autism symptoms are more apparent and easier to recognize in youngsters between 2 and 3 years old. ASD affects one in 68 children in the United States, according to [2]. Existing statistics indicate that approximately 1.5% of the global population is autistic, and it is assumed that a large proportion of autistic people remain unidentified. Consequently, there is a strong need for rapid diagnosis facilities due to the increased recognition of ASD. Traditional diagnostic approaches for ASD involve medical practitioners undertaking a clinical evaluation of the patient's psychological age during clinical assessment. This generally acknowledged method is called Clinical Judgment (CJ). Autism Diagnostic Interview (ADI) and The Autism Diagnostic Observation Schedule (ADOS) are another two conventional diagnostic techniques [3]. A formal diagnosis of autism spectrum disorder is a lengthy procedure since it involves time for learning, undertaking many inquiries, rating, and agreement coding [4].

Recently, the machine learning (ML) technique has emerged as a viable alternative for autism spectrum disorder (ASD) diagnosis because it offers various advantages, including a reduction in diagnostic time, an increase in ASD high detection, and the identification of influential factors. In machine learning, neural networks [5] for deductive reasoning can present difficulties, especially if the model makes medical-associated decisions. The input data quality can have a significant impact on the performance of a neural network. When designing a neural network to tackle a health-related problem, it is crucial to carefully consider the data's quality and characteristics. Our investigation into autism spectrum

disorder yielded unbalanced results. Consequently, we used the Synthetic Minority Oversampling Technique (SMOTE) to balance our autism dataset.

Several studies have attempted to diagnose and treat ASD employing a variety of ML techniques. Bala et al. [6] presented a machine learning system to better identify ASD across age groups. For this, different classification techniques were used on these datasets. SVM outperformed other classifiers for autism spectrum datasets. Lastly, Shapley Additive Explanations (SHAPs) were employed to determine the most accurate feature sets. Hasan et al. [7] demonstrated an effective methodology for evaluating Machine Learning (ML) strategies for early ASD detection. This system used four Attribute Scaling (AS) techniques and eight basic but efficient ML algorithms to classify feature-scaled datasets. AB and LDA identified Autism with the best accuracy of 99.25% and 97.95% for toddlers and children, and 99.03% and 97.12% for adult and teenage datasets, respectively. In another work, Rodrigues et al. [8] employed machine learning and functional magnetic resonance imaging to find potential indicators of ASD prevalence. They utilized the ADOS score as a measure of severity. By achieving 73.8% accuracy on cingulum regions, their findings indicate a functional distinction amongst ASD subclasses. A framework was proposed by Raj et al. [9] in which multiple machine learning algorithms are implemented. The accuracy of CNN-based predicting models for ASD screening in data for adults, adolescents, and children was 99.53%, 96.68%, and 98.30%, respectively.

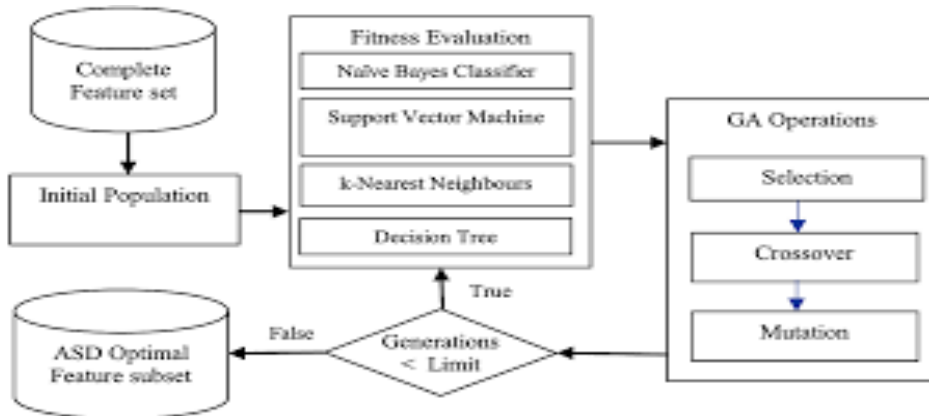


Fig 1: Optimization of Machine Learning Models

With the objective of improving diagnosis, Hossain et al. [10] attempted to identify the most important characteristics and automate the early diagnosis employing existing classification algorithms. Sequential minimum optimization (SMO)-based SVM is superior to all other ML techniques with regard to accuracy, according to their observations. They showed that the Relief Qualities method is the most effective at identifying the most important features in ASD datasets. Akter et al. [11] collected initially identified Autism datasets pertaining to toddlers, children, adolescents, and adults, and employed a variety of feature transformation techniques to these datasets. The performance of several classification approaches was then evaluated using these altered ASD datasets. SVM performed best on the toddler dataset, whereas Adaboost performed best on the children dataset, GImboost on the teenage dataset, and Adaboost on the adult dataset. They identified key traits that are highly predictive for ASD and obtained 98.77%98.77% accuracy. Furthermore, Thabtah et al. [4] suggested a new architecture for adult and adolescent autism screening based on machine learning that includes essential aspects and demonstrates predictive analysis with logistic regression to uncover key knowledge on autism screening. In addition, they ran a comprehensive feature investigation on the different datasets using Chi-square testing (CHI) and Information Gain (IG) to discover the influential characteristics. The acquired results indicate that the ML technique was able to produce forecasting systems with satisfactory performance.

II. RELATED WORK

Autism spectrum disorders (ASD) refer to a group of complex neurodevelopmental disorders of the brain such as autism, childhood disintegrative disorders, and Asperger’s syndrome, which, as the term “spectrum” implies, have a wide range of symptoms and levels of severity [1]. These disorders are currently included in the International Statistical Classification of Diseases and Related Health Problems under Mental and Behavioral Disorders, in the category of

Pervasive Developmental Disorders [2]. The earliest symptoms of ASD often appear within the first year of life [3–6] and may include lack of eye contact, lack of response to name calling, and indifference to caregivers. A small number of children appear to develop normally in the first year, and then show signs of autism between 18 and 24 months of age [5], including limited and repetitive patterns of behavior, a narrow range of interests and activities, and weak language skills. As these disorders also affect how a person perceives and socializes with others, children may suddenly become introverted or aggressive in the first five years of life as they experience difficulties in interacting and communicating with society. While ASD appears in childhood, it tends to persist into adolescence and adulthood [7].

Advanced information technology that uses artificial intelligence (AI) models has helped to diagnose ASD early through facial pattern recognition. Yolcu et al. [8] used the convolutional neural network (CNN) algorithm to train data for extracting components of human facial expressions and proposed the use of such algorithm to detect facial expressions in many neurological disorders [9]. In 2018, Haque and Valles [10], using deep learning approaches, updated the Facial Expression Recognition 2013 dataset to recognize facial expressions of autistic children. Rudovic et al. [1] presented the CultureNet deep learning model that was used to identify 30 videos.

Several studies have discovered important features of autism in numerous ways to detect autism, such as feature extraction [2], eye tracking [3], facial recognition, medical image analysis [9], and voice recognition [4]. However, facial recognition plays a more significant role in detecting autism than the person's emotional state. Facial recognition is a common way to identify persons and to prove that they are normal or abnormal. It involves mining pertinent information to disclose behavior patterns [5, 6]. Duda et al. [7] introduced a new method of producing samples of differentials between autism and attention deficit hyperactivity disorder (ADHD) and using such differentials to recognize autism. Sixty five samples of differentials on facial expressions of social responsiveness were collected for the datasets. Deshpande et al. [12] designed metrics for studying brain activities to identify autism. AI and soft computing approaches have also been applied to detect autism. Different studies have been conducted on detecting autism, but few of them have focused on brain neuroimaging. Parikh et al. [11] used machine learning approaches to develop a system for extracting the characteristics of autism. They gathered their dataset from 851 persons whom they classified as having and not having ASD. Thabtah and Peebles [10] used rule-based machine learning (RBML) to detect ASD traits. Al Banna et al. [12] developed a smart system for monitoring ASD patients during the coronavirus disease 2019 (COVID-19) pandemic.

Many real-life applications of AI and machine algorithms have been designed to help solve social problems. AI has been used in all health applications to help doctors control diseases such as autism. Artificial neural networks have been particularly focused on as a way to extract features of ASD patients that can be used to discriminate between persons with and without ASD. Among the techniques that have been used or proposed to detect autism in children are deep learning techniques, particularly, CNNs and recurrent neural networks (RNNs), and the bidirectional long short-term memory (BLSTM) model. Lately, more studies have been conducted to diagnose ASD using machine learning approaches such as, brain imaging, analysis of data on physical biomarkers, assessment of the behavior of persons with autism, and assessment of clinical data using the machine learning approach.

III. METHODS

The convolutional neural network (CNN) is one of the most famous deep learning algorithms. It takes the input image and assigns importance to learnable weights and biases in order to recognize the class of the image. The neuron can be said to be a simulation of the communication pattern of the neurons of the human brain through the interconnection and communication between cells. In this section, we will explain the basic components of the CNN model: the input layer, convolutional layer, activating function, pooling layer, fully connected layer, and output prediction.

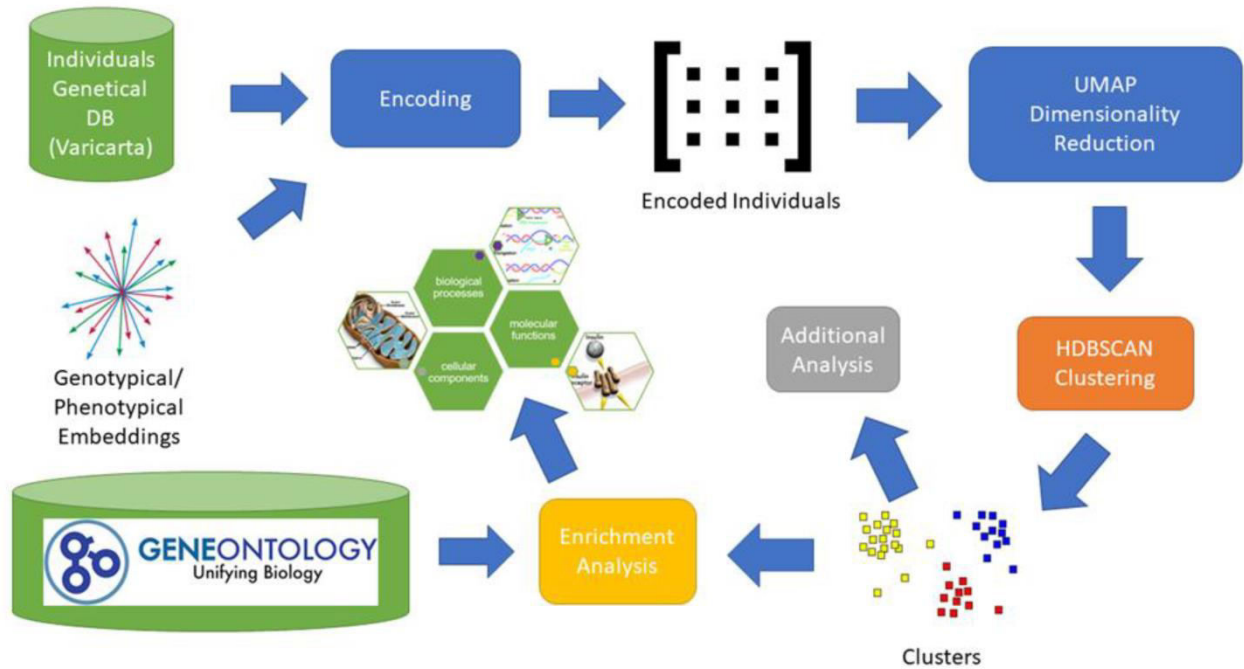


Fig 2: Endophenotypes in Autism Spectrum

The input of the convolutional layer is an image as a matrix of pixel values. The objective of the convolutional layer is to reduce the images into a form that can be easily processed without losing their important features that will help to detect autism. The first layer of the CNN model is responsible for extracting low-level features such as edges and color. The build of the CNN model allowed us to add more layers to it to enable it to extract the high-level features that will help it understand images. Due to the large number of parameters outputted from the convolution layer, which may significantly prolong the arithmetic operations of the matrices, the number of weights was reduced by using one of the following two techniques: max pooling or average pooling. Max pooling is based on the maximum values in each window in the stride, while average pooling is based on the mean value of each window in the stride.

IV. RESULT ANALYSIS

Several studies have been conducted utilizing ASD datasets, although ASD prediction remains in need of substantial improvement. In our research, we collected ASD datasets and balanced them using the SMOTE method. Then, we applied Decision Tree, Naive Bayes, K-Nearest Neighbors, Random Forest, Gradient Boosting, Extreme Gradient Boosting, Multilayer Perceptron, Support Vector Machine, Ada Boost, and Logistic Regression classifiers after transforming the features using various feature transformed techniques, including Standard Scalar, Unit Vector Scalar, Robust Scalar, and Yeo Johnson transformation. All classifiers performed well utilizing the standard scalar approach, with AB achieving the best performance. Subsequently, we applied the feature selection techniques Boruta, Correlation, Mutual Information, and Recursive Feature Elimination, and we achieved good results using the RFE technique. We decreased execution time and memory requirements by utilizing FS to locate ASD as rapidly as feasible. Finally, the grid search hyperparameter optimization technique was implemented. Here, we found that all classifiers improved their performance, with AB, LR, SVM, and MLP producing the best results.

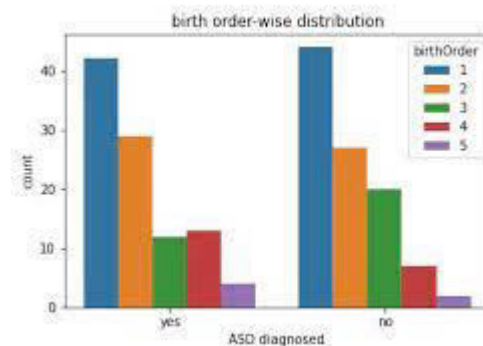


Fig 3: Predictive Analysis of Autism Spectrum

Our results suggest a variety of critical and pertinent features for early diagnosis of ASD. The most important features are A6, A9, A2, A8, and A4 dependent on the log-based relationship. The most important characteristics of an ML model are A8, A7, A6, A1, and A2. In addition, we found significant indicators such as A6, A2, and A8 that are identical for both log-based association and ML methods. Our investigation suggests that key attributes are sufficient for recognizing ASD, which will facilitates the effective application of ASD diagnosis.

The suggested model is contrasted with relevant prior findings in **Table 7**. The majority of previous research utilized version-1 ASD datasets, whereas a few studies utilized version-2. Akter et al. utilized the technique of feature transformation and obtained the maximum result accuracy (98.77%), KS (97.1), AUROC (99.98%), and Log Loss (3.01). In a different study, Bala et al. used the feature selection technique and achieved the top results for accuracy (97.82%), KS (94.87), AUROCO (99.7%), and F1 Score (97.8). Hasan et al. then performed a feature transformation strategy that yielded the highest accuracy (99.25%), KS (98.97), precision (99.89%), recall (98.45%), F1 Score (99.1%), and Log Loss (0.0802).

V. CONCLUSIONS

In our work, the proposed ML architecture was used to yield more precise and effective results for the rapid diagnosis of ASD. We applied FT techniques to the ASD samples, examined the modified dataset using many classifiers, and evaluated their efficacy. Next, we employed feature extraction strategies to yield fewer characteristics from ASD screening methodologies while preserving performance consistency. In addition, we utilized the hyperparameter optimization method to enhance the performance of each classifier. Our research reveals that the standard scalar feature transformation and RFE feature selection techniques are superior to others.

Our findings may aid in the identification of ASD characteristics, making it easier for patients and their families to gain the necessary support to improve their physical, social, and academic health. This study's weaknesses include the absence of specific details such as recordings and images of cases and controls. In the future, we will implement deep learning algorithms that enable the discovery of novel, non-conventional ASD traits from a complex set of characteristics. In addition, future research could examine cluster analysis to identify endophenotypes, evaluate the role of development in facilitating evaluation, and improve diagnosis and therapy.

REFERENCES

1. Crane, L.; Batty, R.; Adeyinka, H.; Goddard, L.; Henry, L.A.; Hill, E.L. Autism diagnosis in the United Kingdom: Perspectives of autistic adults, parents and professionals. *J. Autism Dev. Disord.* **2018**, *48*, 3761–3772. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
2. Thabtah, F.; Spencer, R.; Abdelhamid, N.; Kamalov, F.; Wentzel, C.; Ye, Y.; Dayara, T. Autism screening: An unsupervised machine learning approach. *Health Inf. Sci. Syst.* **2022**, *10*, 26. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
3. Thabtah, F.; Kamalov, F.; Rajab, K. A new computational intelligence approach to detect autistic features for autism screening. *Int. J. Med. Inform.* **2018**, *117*, 112–124. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]

4. Thabtah, F.; Abdelhamid, N.; Peebles, D. A machine learning autism classification based on logistic regression analysis. *Health Inf. Sci. Syst.* **2019**, *7*, 12. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
5. Rocchetti, M.; Delnevo, G.; Casini, L.; Mirri, S. An alternative approach to dimension reduction for pareto distributed data: A case study. *J. Big Data* **2021**, *8*, 39. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
6. Bala, M.; Ali, M.H.; Satu, M.S.; Hasan, K.F.; Moni, M.A. Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder. *Algorithms* **2022**, *15*, 166. [[Google Scholar](#)] [[CrossRef](#)]
7. Hasan, S.M.; Uddin, M.P.; Al Mamun, M.; Sharif, M.I.; Ulhaq, A.; Krishnamoorthy, G. A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders. *IEEE Access* **2022**, *11*, 15038–15057. [[Google Scholar](#)] [[CrossRef](#)]
8. Rodrigues, I.D.; de Carvalho, E.A.; Santana, C.P.; Bastos, G.S. Machine Learning and rs-fMRI to Identify Potential Brain Regions Associated with Autism Severity. *Algorithms* **2022**, *15*, 195. [[Google Scholar](#)] [[CrossRef](#)]
9. Raj, S.; Masood, S. Analysis and detection of autism spectrum disorder using machine learning techniques. *Procedia Comput. Sci.* **2020**, *167*, 994–1004. [[Google Scholar](#)] [[CrossRef](#)]
10. Hossain, M.D.; Kabir, M.A.; Anwar, A.; Islam, M.Z. Detecting autism spectrum disorder using machine learning techniques. *Health Inf. Sci. Syst.* **2021**, *9*, 386. [[Google Scholar](#)] [[CrossRef](#)]
11. Akter, T.; Shahriare Satu, M.; Khan, M.I.; Ali, M.H.; Uddin, S.; Lió, P.; Quinn, J.M.W.; Moni, M.A. Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders. *IEEE Access* **2019**, *7*, 166509–166527. [[Google Scholar](#)] [[CrossRef](#)]
12. Pietrucci, D.; Teofani, A.; Milanese, M.; Fosso, B.; Putignani, L.; Messina, F.; Pesole, G.; Desideri, A.; Chillemi, G. Machine Learning Data Analysis Highlights the Role of Parasutterella and Alloprevotella in Autism Spectrum Disorders. *Biomedicines* **2022**, *10*, 2028. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
13. Omar, K.S.; Mondal, P.; Khan, N.S.; Rizvi, M.R.K.; Islam, M.N. A Machine Learning Approach to Predict Autism Spectrum Disorder. In Proceedings of the 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, 7–9 February 2019; pp. 1–6. [[Google Scholar](#)] [[CrossRef](#)]



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details