Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 2, No.5 : 2024
ISSN : **1906-9685**

MACHINE LEARNING FOR THE EARLY IDENTIFICATION OF AUTISM SPECTRUM DISORDER IN TODDLERS

Dr.Banoth Samya, Associate Professor, Department of CSE, CVR College of Engineering, Hyderabad, India : samyabanoth@gmail.com
Dr. Srikanth Lakumarapu, Associate Professor, Department of CSE, CVR College of Engineering, Hyderabad, India : dr.srikanthcse@gmail.com
Dr. T. Venkata Ramana, Associate Professor, Department of CSE – AIML, CVR College of Engineering, Hyderabad, India : meetramana1204@gmail.com

Abstract
Autism Spectrum Disorder (ASD), also known as "autism," is a psychiatric disorder that impacts an individual's language, thinking, and social skills. Autism Spectrum Disorder (ASD) is a widespread condition, affecting around 1 in 54 births and around 1% of the world's population. Regrettably, despite its widespread occurrence, the etiology and remedy for autism remain elusive, presenting substantial difficulties for parents who fear their kid may have ASD. Timely identification of autism is essential for a child's progress, but it can be exceedingly challenging when symptoms emerge during the child's maturation. Diagnostic tests administered to children aged 2 to 3 years are generally less dependable compared to those conducted on children aged 4 to 5 years. This condition is concerning since timely diagnosis is crucial for autistic individuals to achieve their developmental objectives effectively. Autism is frequently distinguished by challenges in social interaction and communication, which can make accurate diagnosis difficult, even with sophisticated instruments such as the ADOS and ADI. This study aims to solve the issues related to the diagnosis of autism by specifically concentrating on enhancing the diagnostic process. The process entails training and evaluating machine learning models, namely Random Forest with Standard Scaler, using a dataset on autism spectrum disorder. The objective is to discover the most influential factors that indicate autism in toddlers. The objective is to provide a quantitative methodology to assist in the early detection and subsequent management of autism, as prompt intervention can alleviate the long-term symptoms associated with the condition. This study intends to utilize machine learning techniques to get significant insights in order to improve the accuracy of diagnosing autism and enhance assistance for individuals with ASD and their families.

Keywords: Autism spectrum disorder, Predictive analytics, Data analysis, Supervised learning.

1. Introduction

ASD is a developmental disorder that involves persistent challenges in social interaction, speech and nonverbal communication, and restricted and repetitive behaviours. In the USA, the prevalence of ASD has increased substantially in the past two decades, with an estimate of every 1 in 44 children to be identified with ASD by age 8 in 2016 [1]. Although there exist evidence-based interventions which improve core symptoms in children with ASD, many children with ASD still experience long-term challenges with daily life, education and employment [2]. Early diagnosis is the key to early intervention for improving the long-term outcomes of children with ASD. However, despite the growing evidence shows that accurate and stable diagnoses can be made by 2 years, in real-world settings, the median age of ASD diagnosis is 50 months. To improve early diagnosis, the American Academy of Paediatrics (AAP) has recommended universal screening among all children at 18-month and 24-month well-child visits in the primary care settings using the Modified Checklist for Autism in Toddlers (M-CHAT) [3], a questionnaire that assesses

children's behaviour for toddlers. However, growing evidence has shown that using M-CHAT alone may not yield sufficient accuracy in detectingASD cases, with a sensitivity below 40% and a positive predictive value (PPV) under 20% [4, 5]. In addition to ASD-specific behavioural questionnaires, general clinical and healthcare records may also contain meaningful signals to differentiate the ASD risks among very young children. Studies have found that children with ASD are oftentimes accompanied by certain symptoms and medical issues
such as gastrointestinal problems, infections and feeding problems. This implies that past diagnosis and healthcare encounter information, commonly available from health insurance claims or Electronic Healthcare Record (EHR), could potentially be used for ASD risk prediction. In fact, medical claims and EHR data have been widely used in the health informatics literature for identifying disease- specific early phenotypes even before the hallmark symptoms start to manifest, such as for chronic diseases like heart failures, diabetes and Alzheimer's disease [6].

2. Literature Survey

This section explains previous studies that use machine learning-based approaches to detect and predict the autism spectrum disorder. The main motive is to analyze and find some limitations to propose a new, better, and improved machine-learning based approach for autism spectrum disorder prediction. Automated algorithms for disease detection are being deeply studied for usage in healthcare. Graph theory and machine learning algorithms were used. For each age range being examined, the pipeline automatically selected 10 biomarkers. In discriminating between ASD and HC,measures of centrality are the most operational [8]. The study [9] used a neural network-based feature selection method from teacher-student which was suggested to have the most discriminating features and applied different classification algorithms. The results are compared with the already presented methods at the overall and site level. The authors in [10] also utilize the neural network to acquire the distributions of PCD for the classification of ASD as it has far more hyper parameters that make the model extra versatile. Payabvash et al. [11] used computer leaning algorithms to classify children with autism based on tissue connectivity metrics, hence, observed decreased connectome edge density in the longitudinal white matter tracts. It illustrated the viability of it in identifying children with ASD, connectome-based machine-learning algorithms.
The authors in [12] conclude that the data may be used to establish diagnostic biomarkers for the progression of autism spectrum disorders and to distinguish those with the condition in the general population. Wang et al. [13] proposed an ASD identification approach which focuses on multi-atlas deep feature representation and ensemble learning technique. In study [14], the multimodal automated disease classification system uses two types of activation maps to predict whether the person is healthy or has autism. It was able to achieve 74% accuracy. Rakić et al. [15] suggested a technique which is based on a system composed of autoencoders and multilayer perceptron. Because of a multimodal approach that included a set of structural and functional data classification classifiers, the highest classification precision was 85.06%. In study [16], advanced deep-learning algorithms are proposed where HPC solutions can increase the accuracy and time of broad fMRI data analysis significantly. Thomas et al. [17] introduced a novel analysis technique to identify changes in population dynamics in functional networks under ASD. They have also introduced machine learning algorithms to predict the class of patients with ASD and normal controls by using only population trend quality metrics as functions. The limitation of this approach is that the outcomes of the classification are highly dependent on the threshold parameter T. Another problem is that despite age variations in the experimental samples, the same spatial normalization design was used for all subjects. The authors in ref [18] proposed a collection of new features based on MRI images using machine learning algorithms to diagnose ASD which achieved 77.7% accuracy using the LDA approach.

3. Results and Discussion

This dataset contains information about various attributes of individuals, including demographic information (age, gender, ethnicity), medical history (jaundice), family history of ASD, and assessment scores (Q-CHAT-10) used to predict the presence of ASD traits. The "Class/ASD Traits" column appears to be the target variable used for classification purposes, indicating whether theindividual exhibits ASD traits or not.

	Case_No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Age_Mons	Qchat-10-Score	Sex	Ethnicity	Jaundice	Family_mem_with_ASD	Who completed the test	Class/ASD Traits
0	1	0	0	0	0	0	0	1	1	0	1	28	3	f	middle eastern	yes	no	family member	No
1	2	1	1	0	0	0	1	1	0	0	0	36	4	m	White European	yes	no	family member	Yes
2	3	1	0	0	0	0	0	1	1	0	1	36	4	m	middle eastern	yes	no	family member	Yes
3	4	1	1	1	1	1	1	1	1	1	1	24	10	m	Hispanic	no	no	family member	Yes
4	5	1	1	0	1	1	1	1	1	1	1	20	9	f	White European	no	yes	family member	Yes
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1049	1050	0	0	0	0	0	0	0	0	0	1	24	1	f	White European	no	yes	family member	No
1050	1051	0	0	1	1	1	0	1	0	1	0	12	5	m	black	yes	no	family member	Yes
1051	1052	1	0	1	1	1	1	1	1	1	1	18	9	m	middle eastern	yes	no	family member	Yes
1052	1053	1	0	0	0	0	0	0	1	0	1	19	3	m	White European	no	yes	family member	No
1053	1054	1	1	0	0	1	1	0	1	1	0	24	6	m	asian	yes	yes	family member	Yes

1054 rows × 19 columns

Figure 2: Sample dataset used for classification of ASD.

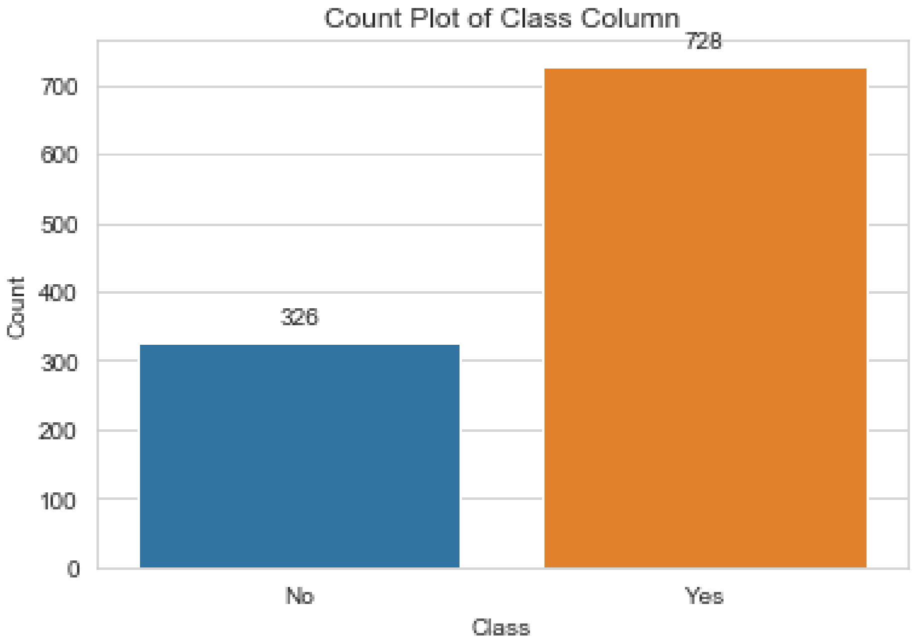

Figure 3: Count plot of class column i.e., Autism or Normal.

```
Accuracy: 100.00
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        68
           1       1.00      1.00      1.00       143

    accuracy                           1.00       211
   macro avg       1.00      1.00      1.00       211
weighted avg       1.00      1.00      1.00       211
```

Figure 4: Obtained accuracy and classification report using XGBoost model.

4. Conclusion

In the analysis of autism prediction in toddlers, the XGBoost model emerges as the superior choice compared to the Naive Bayes model. This assessment is based on various performance metrics and evaluations. Firstly, in terms of accuracy, the XGBoost model demonstrates a notably higher accuracy score when compared to the Naive Bayes model. This suggests that XGBoost excels at correctly classifying cases, a crucial aspect of any predictive model. Secondly, a closer examination of the classification report reinforces the superiority of the XGBoost model. The classification report provides insights into precision, recall, and F1-score values for both classes—likely autism and not autism. The XGBoost model consistently showcases higher values across these metrics, indicating a more balanced and accurate classification of positive and negative cases. Furthermore, the confusion matrix, which offers a detailed breakdown of true positives, true negatives, false positives, and false negatives, reflects better performance by the XGBoost model.

References

[1] Maenner MJ, Shaw KA, Bakian AV, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 Years - autism and developmental disabilities monitoring network, 11 Sites, United States, 2018. MMWR Surveill Summ 2021;70:1–16.

[2] McPheeters ML, Weitlauf A, Vehorn A. U.S. preventive services Task force evidence syntheses, formerly systematic evidence reviews. screening for autism spectrum disorder in young children: a systematic evidence review for the US preventive services Task force. Rockville (MD): Agency for Healthcare Research and Quality (US), 2016

[3] Lipkin PH, Macias MM, Council on children with disabilities, section on developmental and behavioral pediatrics. Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. Pediatrics 2020;145. doi:10.1542/peds.2019-3449.

[4] Guthrie W, Wallis K, Bennett A, et al. Accuracy of autism screening in a large pediatric network. Pediatrics 2019;144.

[5] Carbone PS, Campbell K, Wilkes J, et al. Primary care autism screening and later autism diagnosis. Pediatrics 2020;146. doi:10.1542/peds.2019-2314.

[6] Park JH, Cho HE, Kim JH, et al. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. NPJ Digit Med 2020;3:46.

[7] Downs J, Velupillai S, George G, et al. Detection of suicidality in adolescents with autism spectrum disorders: developing a natural language processing approach for use in electronic health records. AMIA Annu Symp Proc 2017;2017:641–9

[8] A. Kazeminejad and R. C. Sotero, “Topological properties of resting-state fMRI functional networks improve machine learning-based autism classification,” Frontiers in Neuroscience, vol. 12, p. 1018, 2019.

[9] N. A. Khan, S. A. Waheeb, A. Riaz, and X. Shang, “A three-stage teacher, student neural networks and sequential feed forward selection-based feature selection approach for the classification of autism spectrum disorder,” Brain Sciences, vol. 10, no. 10, p. 754, 2020.

[10] M. N. Parikh, H. Li, and L. He, “Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data,” Frontiers in Computational Neuroscience, vol. 13, no. 9, p. 9, 2019.

[11] S. Payabvash, E. M. Palacios, J. P. Owen et al., “White matter connectome edge density in children with autism spectrum disorders: potential imaging biomarkers using machine-learning models,” Brain Connectivity, vol. 9, no. 2, pp. 209–220, 2019.

[12] R. M. Thomas, S. Gallo, L. Cerliani, P. Zhutovsky, A. El-Gazzar, and G. van Wingen, “Classifying autism spectrum disorder using the temporal statistics of resting-state functional MRI data with 3D convolutional neural networks,” Frontiers in Psychiatry, vol. 11, p. 440, 2020.

[13] Y. Wang, J. Wang, F.-X. Wu, R. Hayrat, and J. Liu, “AIMAFE: autism spectrum disorder identification with multi-atlas deep feature representation and ensemble learning,” Journal of Neuroscience Methods, vol. 343, Article ID 108840, 2020.

[14] M. Tang, P. Kumar, H. Chen, and A. Shrivastava, “Deep multimodal learning for the diagnosisof autism spectrum disorder,” Journal of Imaging, vol. 6, no. 6, p. 47, 2020.

[15] M. Rakić, M. Cabezas, K. Kushibar, A. Oliver, and X. Lladó, “Improving the detection of

autism spectrum disorder by combining structural and functional MRI information," NeuroImage: Clinic, vol. 25, Article ID 102181, 2020.
[16] T. Eslami, J. S. Raiker, and F. Saeed, "Explainable and Scalable Machine-Learning Algorithms for Detection of Autism Spectrum Disorder Using fMRI Data," 2020, https://arxiv.org/abs/2003.01541.
[17] E. E. Thomas Martial, L. Hu, and S. Yuqing, "Characterising and predicting autism spectrum disorder by performing resting-state functional network community pattern analysis," Frontiers in Human Neuroscience, vol. 13, p. 203, 2019.
[18] S. Mostafa, L. Tang, and F.-X. Wu, "Diagnosis of autism spectrum disorder based on eigenvalues of brain networks," IEEE Access, vol. 7, pp. 128474–128486, 2019.