MACHINE LEARNING ALGORITHMS FOR THE EARLY DIAGNOSIS OF AUTISM SPECTRUM DISORDER IN CHILDHOOD

Ashlin Femi R N[1], Anto Kumar[2]

[1]Assistant Professor, Dept. of CSE, Narayanaguru College of Engineering, Manjalumoodu, Kanyakumari District
[2]Professor, St. Xavier's Catholic College of Engineering, Nagercoil, Kanyakumari District.

Abstract - *Receiving an early diagnosis for individuals with autism spectrum disorder (ASD) can make a major difference in their behaviour, abilities, and language development. This article provides a concise and representative overview of artificial intelligence's current involvement in autism evaluation. Many researchers believe that artificial intelligence plays a key role in the early identification of autism since it aids physicians in completing diagnoses faster and with more reliable findings. This study highlights the use of smart technology in the autism diagnosis process by presenting several artificial intelligence applications that are now in use or are in the early stages of development. An early and precise diagnosis is critical for a tailored and effective intervention that supports the child's academic and personal growth. Clinical standardised testing are currently the sole techniques of diagnosis for ASD. This not only necessitates more time for diagnosis, but it also results in a significant rise in medical costs. Machine learning approaches are being utilised in addition to traditional methods to increase the precision and time required for diagnosis.*

Key Words: Autism Spectrum Disorder, ASD, Machine learning, classification, early diagnosis

1. INTRODUCTION

Screening instruments, interviews, and clinical observations are all used to diagnose autistic, which is a neurological disease. It is mostly focused on the examination of behavioural patterns via questionnaires completed by parents or guardians. ADOS is a semistructured test of speech, social interaction, and imagination that is aimed to identify children on the autism spectrum [1]. Autism is a widespread developmental condition because it affects a person at all stages of their life. The term "pervasive" denotes that the disorder has an overall impact on a person's development [2], while "disorder" denotes a sense of separation from the norm. The fact that autism is a congenital illness does not mean that symptoms appear right away. They are usually noticeable before entering primary school. The chances of a therapeutic intervention have been considerably lowered for children who were not diagnosed before the age of eight years. The idea is to use the assessment tools and classifiers early in the process [3]. (Hierarchical Agglomerative Clustering, Gaussian Mixture Models). Their goal was to utilise these groupings to tailor therapies to each individual. The findings of this study show that machine learning can distinguish autism phenotypes [12]. Heinsfeld et al. looked at how deep learning algorithms may be used to identify autistic people based on brain activity patterns. They employed the ABIDE (Autism Brain Imaging Data Exchange) database, and the algorithms had a 70 percent accuracy rate. As a consequence, the high percentage demonstrates that machine learning approaches have a lot to offer and are a potential tool for assessing mental problems in general [13].

Regardless of the fact that autism spectrum disorders (ASD) are thought to be neurological, no brain biomarkers have been discovered, and diagnosis is still reliant on behavioural factors. With this in mind, Chen et al. chose 252 functional MRI (MRI) images (low head motion) from the Autism Brain Imaging Data Exchange (ABIDE), which included Typically Developing (TD) and ASD patients (n = 126 each) who were matched for non-verbal IQ, head motion, and age [14]. While random forest (RF) obtained a high accuracy of 91 percent for diagnostic classification, support vector machines in conjunction with particle swarm optimization and recursive feature reduction fared moderately (with accuracies for validation datasets of 70 percent) [14]. Kosmicki et al. employed machine learning to see if algorithms could categorise people into two groups based on whether or not they were on the autism spectrum [15]. We see that these abridged classifiers preserve the diagnostic validity of the original algorithm, and that if a reduced number of behaviours are examined using machine learning methods, high percentages of validity at the autism prognosis may be achieved [15]. In order to increase the accuracy and quality of prediction, Vaishali and Sasikala looked into a machine learning repository that used swarm intelligence. They show in their study that ten database features can distinguish between those who are on the spectrum and those who aren't, and that this strategy is accurate to the tune of 97, 95 percent [16].

In order to create a subject-transfer decoder, Koyamada et al. (2015) looked into a deep learning model (DNN) model. The authors created a decoder for viewing distinct aspects of all people in the dataset via principle sensitivity analysis (PSA). The two hidden layers in the centre categorise brain events into seven human categories from 499 people in their proposed neural net, which has two hidden layers and a softmax output layer. ASD has been found to impact global brain networks by disrupting functional connections between numerous brain areas. As a result, several research works are working to categorise ASD and control participants depending on the operational connectivity patterns in the brain.

Machine learning approaches have been used to supplement conventional methods in order to enhance the precision and time required for diagnosis. On our dataset, we used models like Support Vector Machines (SVM), Random Forest Classifier (RFC), Nave Bayes (NB), Logistic Regression (LR), and KNN to create prediction models. The major goal of our article is to see if the kid is at risk for ASD in its early stages, that will speed up the diagnostic procedure. According to our findings, Logistic Regression provides the maximum accuracy for the dataset we chose. Early detection and treatment of Autism Spectrum Disorder are critical since they assist to reduce or alleviate symptoms to some extent, therefore enhancing the individuals personal overall standard of living. Autism Spectrum Quotient (AQ), Childhood Autism Rating Scale (CARS-2), and Screening Tool for Autism in Toddlers and Young Children are some of the screening approaches used to diagnose ASD in children (STAT).

2. METHODOLOGY

The proposed methodology is given in Fig.1. Dr. Fadi Thabtah [18] created the dataset [19], which includes categorical, continuous, and binary features. The dataset has 1054 instances with 18 characteristics when it was first created (including class variable). We start by removing missing values and outliers from the dataset, removing noise, and encoding categorical characteristics. We also use feature engineering to choose the most useful characteristics from among the many available in the data set. This minimises data dimensionality, which improves training speed and efficiency. Once the data set has been preprocessed, classification algorithms like Logistic Regression, Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, and Random Forest Classifiers are used to predict the output label (ASD or no ASD). Each classifier's accuracy is measured and compared.

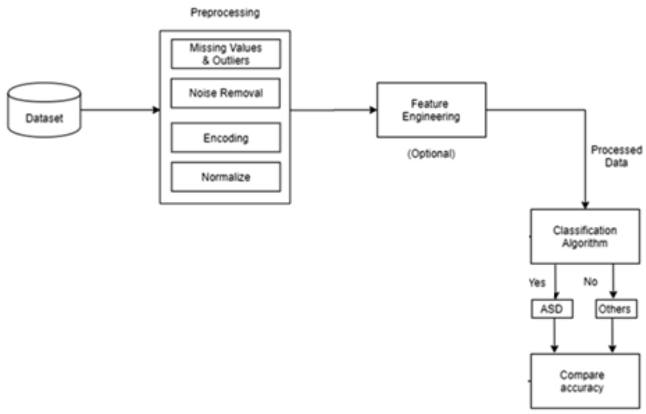

Fig. 1. Proposed methodology

We had to preprocess the data because it included several non-contributing and category features. The adjustments that are made to a data collection before it is fed to the model are referred to as preprocessing. It's performed to clean up raw or noisy data so it's better for training and evaluation. Label encoding is used to handle with categorical data. Label encoding translates labels into numeric form so that they may be read by machines. The very same value is allocated to repeated labels as it was previously. Binary label encoding was chosen for four characteristics with two classes (Sex, Jaundice, Family mem with ASD, and Class/ASD Traits). When there are more than two classes, Label Encoding is useless. One-Hot Encoding is used for multiclass characteristics to prevent the model's hierarchical ordering.

2.1 Classification Algorithms

The dataset [19] was divided into two parts: a training set and a test set. The classification model will be trained using the training set, which contains 80% of the data (843 samples). The remaining 20% of the data (211 samples) will be used to evaluate the model's accuracy and efficacy on previously unknown data, and will be known as the testing data set. This haphazard data division into training and testing sets aids us in determining if our model is overfitting or underfitting. The model is overfitting the data whether it has a low training error but a high testing error. The model, on the other hand, is underfitting the data if it has a large training and testing error. After having performed data preprocessing (4.1), we applied five classification models, namely Logistic Regression, Naive Bayes, Support Vector Machine, K-Nearest Neighbors, and Random Forest Classifier, and compared the performance of each based on accuracy achieved and F1 score (Table 1). For the purpose of evaluating the performance of all these models, we have used the confusion matrix and F1 score. Table 1, shows a comparison of all the classification models we used.

Table 1. A Comparison of the Applied Machine Learning Models

	LR	NB	SVM	KNN	RFC
Accuracy	97.15 %	94.79 %	93.84 %	90.52 %	81.52 %
F1 score	0.98	0.96	0.95	0.93	0.88

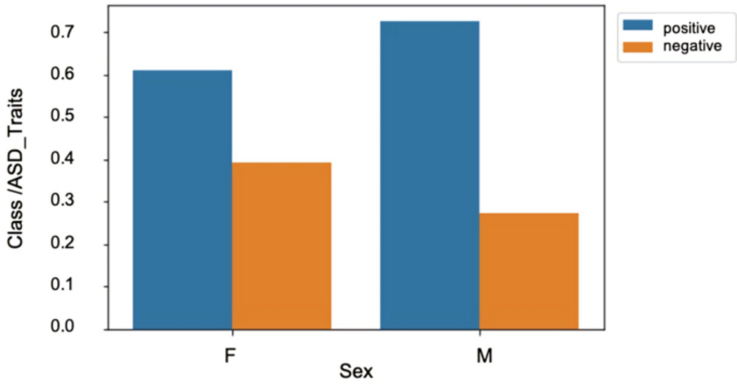

Fig. 3. Gender distribution of ASD traits

3. ANALYSIS

The majority of ASD positive cases in children occur between the ages of 36 and 48 months. Between the ages of 15 and 20, the lowest number of instances were found. Significant indications of autism appear around the age of three years, as seen in the graph (Fig. 2).

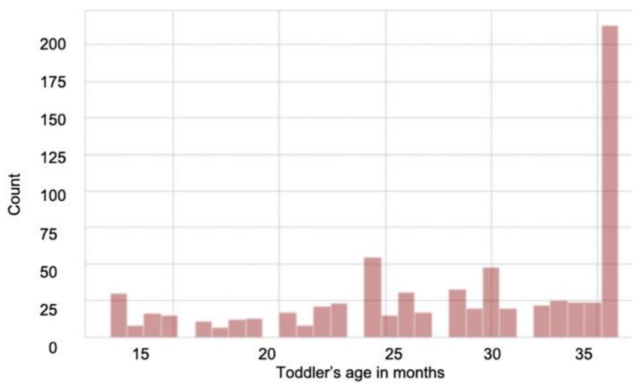

Fig 2: Age distribution of ASD positive

We plotted a gender distribution graph of the ASD traits observed in males and females. It can be concluded that ASD is more prevalent in males than in females as depicted in Fig 3.

4. CONCLUSION AND FUTURE WORK

The evaluation of ASD behavioural features takes time, which is made more difficult by overlapping symptomatology. There is presently no diagnostic test or screening tool that has been specifically created to detect the development of ASD. We created an automated ASD prediction model based on the diagnostic datasets of each participant's minimal behaviour sets. Logistic Regression was found to have the greatest accuracy of the five models we tested on our dataset.

The lack of substantial, open-source ASD datasets is the fundamental constraint of our study. A vast dataset is required to develop an accurate model. We couldn't find enough occurrences in the dataset we utilised. Our study, on the other hand, has contributed to the development of an automated model that can help doctors predict autism in youngsters. To increase generalisation in the future, we will investigate utilising a larger dataset. To increase the system's resilience and overall performance, we propose to use deep learning approaches that combine CNNs with categorization. Overall, our study has led to the examination of numerous classification models capable of reliably detecting ASD in children with certain characteristics based on behavioural and medical data. Other researchers can utilise the study of these classification models as a starting point for additional research into this dataset or other Autism Spectrum Disorder data sets.

REFERENCES

[1] M.E Block, V.E Block, and P. Halliday, “What is autism?,” Teaching Elementary Physical Education, vol. 17, no. 6, pp 7-11, November 2006.

[2] A. Gena, Autism and pervasive developmental disorders. Athens: Idiotiki. 2002.

[3] K. Baraka, F. Melo, and M. Veloso, “Simulating Behaviors of Children with Autism spec-trum disorder through reversal of the autism diagnosis process,” In Proc. of the Portuguese Conference on Artificial Intelligence, 2017, 753-765. https://doi.org/10.10 07/978-3-319-65340-2_61

[4] F. Thabtah, and D. Peebles, “A new machine learning model based on induction of rules for autism detection,” Health Informatics Journal, January 2019.

[5] J. Daniels, J. Schwartz, A. Nikhila, M. Du, and D.P Wall, “The GapMap project: a mobile surveillance system to map diagnosed autism cases and gaps in

autism services globally," Molecular Autism, vol.8, no. 55, October 2017.

[6] 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 for Autism in Toddlers in 1,000 cases and 3,000 controls," Child Adolescent Psychiatry, vol. 51, no. 2, pp. 202-212, February 2012.

[7] Q. Tariq, J. Daniels, J.N. Schwartz, P. Washington, H. Kalantarian, and D.P. Wall, "Mobile detection of autism through machine learning on home video: A development and prospective validation study," PLoSMed, vol. 15, no. 11, November 2018.

[8] J. Kaur and D. Kakkar, "Fuzzy Based Integrated Diagnostic System for Neurodevelopmental Disorders," In Proc. of the 6th International Conference on Signal Processing and Integrated Networks (SPIN), 2019, 132-136.

[9] I. Vlachavas, P. Kefalas, N. Vasiliadis, F. Kokoras, and I. Sakelariou, Artificial Intelligence, Greece: University of Macedonia Press, 2011.

[10] F. Hsiung Hsu, Behind Deep Blue Building The Computer That Defeated The World Chess Champion, Princeton NJ: Princeton University Press, 2002. [E-book] Available: Google Books.

[11] A. Sharma, and A. Sharma, "Machine Learning: A Review of Techniques of Machine Learning," JASC: Journal of Applied Science and Computations, vol. 5, no. 7, pp 538-541, December 2018.

[12] E. Stevens, D.R. Dixon, M.N. Novac, D. Granpeesheh, T. Smith, and E. Linsteada, "Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning," International Journal of Medical Informatics, vol. 129, pp. 29-36, May 2019.

[13] A.S Heinsfeld, A.R. Franco, C. Craddock, A. Buchweitz, and F. Meneguzzia, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," Clinical, vol. 17, pp. 16-23, August 2017.

[14] C.P. Chen, C.L. Keown, A. Jahedi, A. Nair, M.E. Pflieger, B.A. Bailey, R.A. Muller, "Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism," NeuroImage: Clinical, vol. 8, pp. 238-245, April 2015.

[15] JA. Kosmicki, V. Sochat, M. Duda, and DP. Wall, "Searching for a minimal set of behaviors for autism detection feature selection-based machine learning," Translational Psychiatry, vol. 5, no. 514, pp. 1-7, February 2015.

[16] R. Vaishali and R. Sasikala, "A machine learning based approach to classify autism with optimum behavior sets," International Journal of Engineering and Technology (IJET), vol. 4, no.4, pp.4216-4219, August 2018.

[17] Baron-Cohen S, Allen J, Gillberg C. Can autism be detected at 18 months? The needle, the haystack, and the CHAT. Br J Psychiatry. 1992;161:839–43.

[18] Thabtah, F. An accessible and efficient autism screening method for behavioural data and predictive analyses. Health Informatics Journal 2019;25(4):1739–55.

[19] Dataset: https:// www. kaggle. com/ fabde lja/ autism- scree ning- fortoddlers. Accessed 1 Oct 2019 g- fortoddlers. Accessed 1 Oct 2019