Early Detection of Autism Spectrum Disorder Using Machine Learning Classification Techniques (EDASD-MLCT)

Asif Mohamed H B[1]*, Dr.Md. Sameeruddin Khan[2] ,Dr. Mohan K G[3], Dr.Parashuram Baraki[4]

[1*]Research Scholar, School of CSE, Presidency University, Bengaluru, Karnataka, India,
[2] Professor & Dean, School of CSE, Presidency University, Bengaluru, Karnataka, India,
3 Professor, Department of CSE, GITAM University, Bengaluru, Karnataka, India.
[4] Professor, Department of CSE , Smt.Kamala and Sri Venkappa M Agadi College of Engineering & Technology, Lakshmeshwar Dist:Gadag
*Corresponding Author: Asif Mohamed H B
*Research Scholar, School of CSE, Presidency University, Bengaluru, Karnataka, India,

Abstract

Autism spectrum disorder (ASD) is a neuro-developmental condition that impairs a person's capacity for speaking, making announcements, and learning new things. Autism can be examined at each level of the age spectrum. Autism sufferers utilise meditation to examine difficult topics, learning disabilities, and other concerns. Autism-related difficulties might affect a person's movement, sense of smell, and mental health. Controlling ASDs requires early discovery of them. The ASD prediction framework supports the behavioural aspect-based analytic model in this experiment without the need for any equipment. The main focus of the ASD prediction approach is the system's utilised child and adolescent analytic model. The behavioural parameters are gathered using the Autism Query collections.

1. Introduction

Although ASD is still in its infancy, it has the potential to significantly affect issues with social interaction, behaviour, and contact. People with ASD frequently exhibit physical characteristics that set them apart from others. Even so, in the approach segment—which was mostly centred on particular people—people with ASD were able to interact with others, hear, execute, and communicate. People with ASD have access to a wide range of possibilities, from talented to more challenging ones, because of their knowledge, concepts, and problem-solving skills. Many conclusions from an investigation of ASD can be applied to research on the syndrome connected to autism. These illnesses are referred to as "ASD." Because it affects how a person perceives and interacts with others, ASD is a mental disease that makes it challenging for a person to interact with others and conduct business. Poor and recurring behavioural habits, as well as physiological, environmental, and hereditary factors, all have an impact on the condition.

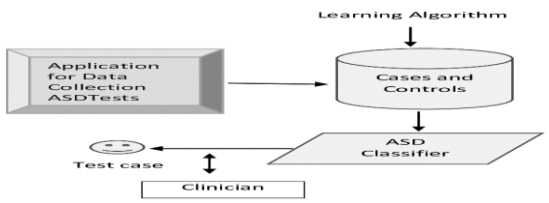

Figure 1. ASD Classifier

Autism affects children more frequently in the US. Every autistic individual has different issues, such as anxiety, seizures, and depression, as can be seen in Figure 1. Autism can start to manifest at the age of two or three and can be identified as early as 18 months. The quality of life is improved by early diagnosis. People with ASD may partake in a variety of activities, connections, and interactions. A person with ASD may be gifted in a variety of pursuits.

For those with ASD, the social cues and restricted behaviours might be challenging. The kids are less focused and engage in different cultural customs. Diagnosis is difficult because there is no valid medical test or blood test to identify ASD. Based on the child's behaviour, doctors may start a diagnosis to enhance their quality of life.

Proposed Methodology

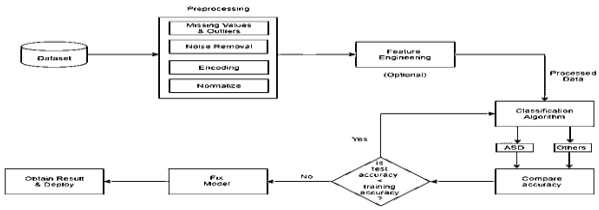

Figure 2 . ASD Prediction

Steps

1. Start
2. Get the data points and the K-value for the number of clusters.
3. Randomly choose the centroids c1, c2,..., ck.
4. Till a predetermined number of iterations have been finished, repeat steps 4 and 5 as necessary.
5. For every data point x I
6. A new centroid is formed for each cluster, j = 1..k, by taking the mean of all the points

doi.org/10.31838/hiv23.03.71 Received: 14.10.22, Revised: 27.11.22, Accepted: 10.12.22.

assigned to that cluster.
7. To begin, find the nearest centroid (c 1, c 2,..c k), then assign the point to that cluster.
8. Stop

Algorithm : Autism Spectrum Disorder – ML Classifiers (EDASD- MLCT)

1. Begin
2. Randomly generate the training and test sets.
3. Explain the model's algorithms in detail.
4. ML Model
5. in the case of (i=0, i13, and i++)
6. mn[i] = model
7. Model.fit();
8. Model.predict();
9. Print the classification report, conflict matrix, and accuracy(i));
10. End

Table 1. ASD Attributes

Several Classification Techniques
Regression Into Reason (LR)

```
A1                    int64
A2                    int64
A3                    int64
A4                    int64
A5                    int64
A6                    int64
A7                    int64
A8                    int64
A9                    int64
A10                   int64
Age_Mons              int64
Qchat-10-Score        int64
Sex                  object
Ethnicity            object
Jaundice             object
Family_mem_with_ASD  object
Class/ASD Traits     object
dtype: object
```

Table 2 – ASD Statistics

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Age Mons	Qchat-10-Score	
count	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000
mean	0.563567	0.448767	0.401328	0.512334	0.524668	0.576850	0.649905	0.459203	0.489564	0.586338	27.867173	5.212524	
std	0.496178	0.497604	0.490400	0.500085	0.499628	0.494293	0.477226	0.498569	0.500128	0.492723	7.980354	2.907304	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	12.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	23.000000	3.000000	
50%	1.000000	0.000000	0.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	1.000000	30.000000	5.000000	
75%	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	36.000000	8.000000	
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	36.000000	10.000000	

The fundamental objective of logistic regression is to identify the model that best describes the association between the relevant binomial feature and a set of independent variables [12].

LDR

Linear discriminant analysis is the most widely used technique for lowering dimensionality. A dataset must be projected onto a lower-dimensional space with sufficient class-separability in order to prevent overfitting (the "curse of dimensionality") and reduce processing costs.

Knn

The foundations of the KNN algorithm are the idea of a distance measure and the presumption that points close to one another are comparable.

Dt

A supervised machine learning technique called a decision tree can be used to categorise or predict data based on the results of previous queries. Due to the nature of supervised learning, the model is trained and tested using data sets that have the necessary classification.

Gaussian NB

Hence, less training data would be required. The biggest issue with NB is that it only works well when a select few features are used. Also, the bias is significant when there is little data.

Svc

The margin is the distance between the nearest training data point and the hyperplane. We used a linear RBF kernel for our initial training and saw that it outperformed a non-linear kernel. Regression and classification are just two of the many applications of the random forest classifier [16]. It necessitates the creation of numerous decision trees in order to be effective for any particular set of data points. Each tree's predictions are collected, and the best solution is then determined by a vote of the general public.

Data Preprocessing

The dataset [3] contains categorical, continuous, and binary variables. Initially, the dataset had 1054 occurrences and 18 attributes (including class variable). Preprocessing refers to the operations carried out on a data collection before giving it to the model. In order to make data more suitable for training and analysis, noisy or unclean data is cleaned.

The attributes Case Number, "Who finished the test," and "Qchat-10-Score," which were not useful, were removed. In order to regulate the category variables, we are utilising label encoding.

Labels must be transformed into numeric form before machines can read them. Repeated labels have a value equal to their corresponding original labels. Label encoding is rendered useless by the addition of any more classes. The "Ethnicity" feature's 11 classes have been one-hot encoded.

Dataset Assessment

The Q-CHAT-10 version is a shortened variant that has been used; it asks a sequence of 10 questions (Table 2). The responses to these queries are converted to binary values based on their class type. These values are assigned during the data gathering process by answering the Q-CHAT-10 questionnaire. The class value "Yes" is given if the Q-CHAT-10 score is more than 3, which denotes the possibility of ASD symptoms. If not, "No" appears in the class result, denoting a lack of ASD traits.

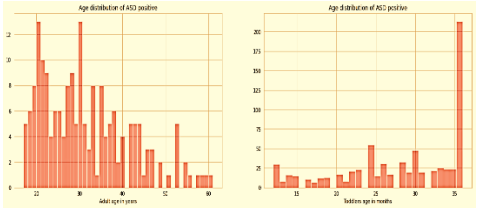

Figure 3 – Age Distribution of ASD Positive in Years and Months

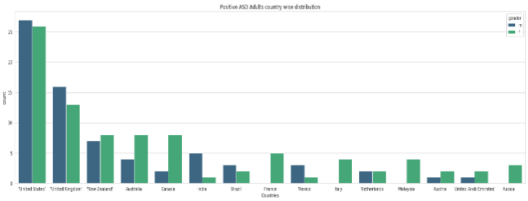

Figure 4 - ASD Country wise Distribution

Evaluation

True positive (TP): As we correctly predicted, the person has ASD. The person does not have autism spectrum disorder (ASD), contrary to our expectation.

False positive (FP): We incorrectly diagnosed ASD even though the patient in question did not have it. This mistake is of kind 1.

False negative (FN): We incorrectly believed that the person did not have ASD, despite the fact that they do. This mistake is of kind 2.

$$precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (1)$$

and recall is defined as :

$$recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (2)$$

Thereforre the F1 score is defined as

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

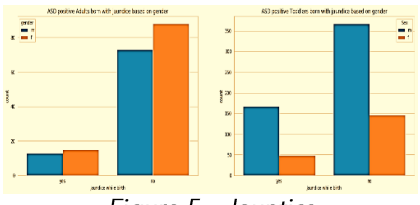

Figure 5 – Jauntics

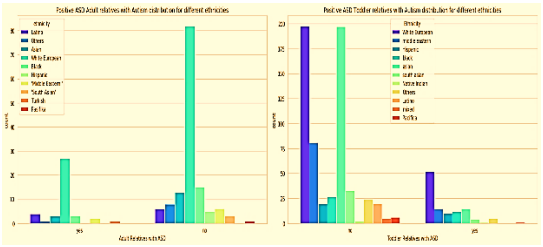

Figure 6- ASD Positive

Table 3 ML Model Comparison		
0	LR	96%
1	KNN	89%
2	NB	96%
3	SVM	91%
4	DT	95%
5	RF	86%

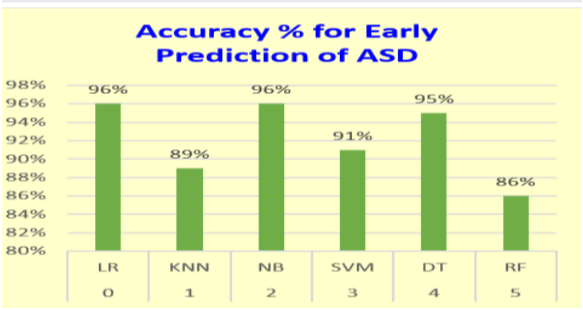

Figure 7. Accuracy

The accuracy score for several classifiers is displayed. Overall, the performance of Logistic Regression is 96% which is better Accuracy when compared to other Machine Learning Approaches.

2. Conclusion

The main focus of the ASD prediction approach is the system's utilised child and adolescent analytic model. The behavioural parameters are gathered using the Autism Query collections. This proposed research work has looked at a number of categorization models that, when combined with the child's behaviour and medical information, can accurately pinpoint ASD in kids who exhibit certain traits.

References

Constantino, John N., Patricia D. Lavesser, Y. I. Zhang, Anna M. Abbacchi, Teddi Gray, and Richard D. Todd. (2007) "Rapid quantitative assessment of autistic social impairment by classroom teachers." Journal of the American Academy of Child & Adolescent Psychiatry 46(12): 1668-1676.

Daniel Bone, Chi-Chun Lee, Matthew P. Black, Marian E. Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan. (2014). "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody." Journal of Speech, Language, and Hearing Research, 57(4): pp.1162-1177.

Daniel Bone, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan. (2015) "Applying machine learning to facilitate autism diagnostics: pitfalls and promises." Journal of autism and developmental disorders

45(5): 1121-1136.
Dennis P. Wall, Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, and Todd F. DeLuca. (2012) "Use of artificial intelligence to shorten the behavioral diagnosis of autism." PloS one, 7(8): e43855.
Dennis Paul Wall, J. Kosmicki, T. F. Deluca, E. Harstad, and Vincent Alfred Fusaro. (2012) "Use of machine learning to shorten observationbased screening and diagnosis of autism." Translational psychiatry, 2(4): e100.
Dr.R.Chinnaiyan, Sabarmathi.G (2017)," Investigations on Big Data Features , Research Challenges and Applications", IEEE International Conference on Intelligent Computing and Control Systems, ICICCS 2017, 782 – 786
Dr.S.Nirmala, R.Chinnaiyan et. al. (2021). Blockchain based Secured Framework for Road Traffic Management using Fog Computing. International Journal of Computational
Intelligence in Control 13(2)
Fadi Thabtah. (2017) "ASD Tests. A mobile app for ASD screening." www.asdtests.com [accessed December 20th, 2017]. [11] Baihua Li, Arjun Sharma, James Meng, Senthil Purushwalkam, and Emma Gowen. (2017) "Applying machine learning to identify autistic adults using imitation: An exploratory study." PloS one, 12(8): e0182652.
Fadi Thabtah. (2017). "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment." In Proceedings of the 1st International Conference on Medical and Health Informatics, pp. 1-6. ACM.
G. Sabarmathi and Dr.R. Chinnaiyan, "Reliable Machine Learning Approach to Predict Patient Satisfaction for Optimal Decision Making and Quality Health Care," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1489-1493.
G. Sabarmathi and Dr.R. Chinnaiyan, "Big Data Analytics Framework for Opinion Mining of Patient Health Care Experience" International Conference on Computing Methodologies and Communication (ICCMC 2020), IEEE Xplore Digital Library
G. Sabarmathi and R. Chinnaiyan, "Reliable feature selection model for evaluating patient home health care services opinion mining systems," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-4, doi: 10.1109/ICAECA52838.2021.9675485.
G.Sabarmathi , Dr.R.Chinnaiyan (2018), "Envisagation and Analysis of Mosquito Borne Fevers – A Health Monitoring System by Envisagative Computing using Big Data Analytics" in ICCBI 2018 – Springer on 19.12.2018 to 20.12.2018 (Recommended for Scopus Indexed Publication IEEE Xplore digital library)
G.Sabarmathi , Dr.R.Chinnaiyan, Reliable Data Mining Tasks and Techniques for Industrial Applications, IAETSD JOURNAL FOR ADVANCED RESEARCH IN APPLIED SCIENCES, VOLUME 4, ISSUE 7, DEC/2017,PP- 138-142, ISSN NO: 2394-8442
G.Sabarmathi, Dr.R.Chinnaiyan "Mining Patient Health Care Service Opinions for Hospital Recommendations" International Journal of Engineering Trends and Technology 69.9(2021):161-167
Hari Pranav A; M. Latha; Ashwin. M. S; R. Chinnaiyan, "BlockchainAs a Service (BaaS) Framework for Government Funded Projects e-Tendering Process Administration and Quality Assurance using Smart Contracts," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-4, doi:10.1109/ICCCI50826.2021.9402348.
Hari Pranav A;M. Senthilmurugan; Pradyumna Rahul K;R. Chinnaiyan , "IoT and Machine Learning based Peer to Peer Platform for Crop Growth and Disease Monitoring System using Blockchain," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402435.
M. Latha, M. Senthilmurugan and R. Chinnaiyan, "Brain Tumor Detection and Classification using Convolution Neural Network Models," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-5, doi:10.1109/ICAE CA52838.2021.9675585.
M. Senthilmurugan, M. Latha and R. Chinnaiyan, "Analysis and Prediction of Tuberculosis using Machine Learning Classifiers," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-4, doi:10.1109/ICAECA52 838.2021.9675482.
Manoj Challa , Dr.R.Chinnaiyan (2019) "Optimized Machine Learning Approach for the Prediction Of Diabetes-Mellitus (Recommended for Scopus Indexed Publication in Springer - Advances in Intelligent Systems and Computing Series)
Preetika B, ; M. Latha; M. Senthilmurugan; R. Chinnaiyan, , "MRI Image based Brain Tumour Segmentation using Machine Learning Classifiers," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-9, doi: 10.1109/ICCCI50826.2021.9402508
R. Chinnaiyan and D. Stalin Alex, "Early Analysis and Prediction of Fetal Abnormalities Using Machine Learning Classifiers," 2021 2nd International Conference on Smart Electronics andCommunication(ICOSEC),2021,pp.1764-1767,doi: 10.1109/ICOSEC51865.2021.9591828. https://ieeexplore.ieee.org/abstract/document/9591828)
R. Chinnaiyan and S. Alex, "Machine Learning Approaches for Early Diagnosis and Prediction of Fetal Abnormalities," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-3, doi: 10.1109/ICCCI50826.2021.9402317.
https://ieeexplore.ieee.org/abstract/document/9402317)

R.Chinnaiyan, Dr.Stalin Alex (2021). Optimized Machine Learning Classifier for Early Prediction of Fetal
Abnormalities . International Journal of Computatio nal Intelligence in Control 13(2)
Vaishali, R., and R. Sasikala. "A machine learning based approach to classify Autism with optimum behaviour sets. (2018) " International Journal of Engineering & Technology 7(4): 18.