Rule Based Classifier for the Detection of Autism in Children

Kusumalatha Karre[1,2](✉) and Dr. Y. Ramadevi[3]

[1] Sreenidhi Institute of Science and Technology, Hyderabad, India
kusumalatha.karre@gmail.com
[2] Osmania University, Hyderabad, India
[3] Chaitanya Bharathi Institute of Technology, Hyderabad, India

Abstract. Autism is a developmental disorder that hinders the life of an autistic child with poor communication and a lack of social skills to carry out their day-to-day work. Detecting autism is very important at an early stage to help the child overcome their learning disabilities. Generally, Autism is diagnosed by specialists in hospitals or therapy centers using procedures that are expensive and time-consuming. Research has been carried out to use various machine learning algorithms to develop intelligent classifiers for autism which can improve accuracy and reduce time. In this paper, we propose a Rule based classifier that generates rules that are combined with machine learning algorithms to detect autism in children by using the QCHAT screening tool. It is the first time Rule based machine learning has been used on a QCHAT screening tool that detects autism during 18–30 months of age. The dataset of QCHAT with rule based classifier has been used for detecting autism and achieved an accuracy of 97.37%. This would be helpful for the doctors and parents to diagnose the child with autism and initiate necessary therapies which can help the child to develop to the fullest.

Keywords: Autism Diagnosis · classification · machine learning · Rule based model

1 Introduction

Autism is a mental disability that hinders the development of a child in various aspects such as communication, gross motor skills, and personal skills. The exact cause of it is not known but it is assumed to be caused due to chemical disturbances in the brain [1]. There is no proper medicine available for it and only intervention helps to develop the necessary skills. As per the Centers for disease control and Prevention (CDC) report on autism, 2020,1 in 54 children has autism [2]. The rest of the paper is organized as follows: Literature Survey, Proposed Rule based classifier model, Results, and Conclusion.

2 Literature Survey

There are various ways of detecting autism and the most common method is to use autism screening and diagnostic tools. The other methods used are functional magnetic resonance imaging (f-MRI) scanning, electroencephalogram (EEG) signals, and Eye

B. Raj et al. (Eds.): ICETE 2023, AER 223, pp. 79–86, 2023.
https://doi.org/10.2991/978-94-6463-252-1_10

gaze. A lot of research has been done to detect autism using f-MRI scanning, EEG signals, and Eye gaze and the accuracy achieved is also good but they cannot be used to identify all the aspects of autism. In this paper, we concentrate on various autism screening tools on which various machine learning algorithms have been used to improve accuracy and reduce the time for identifying children with autism.

Various Screening tools available are:

- Ages and Stages Questionnaires (ASQ) (1 month-5.6 years): It is a questionnaire to be filled out by the parents and has 19 questions related to gross and fine motor skills, communication, and personal adaptive skills [3].
- Communication and Symbolic Behavior Scales (CSBS) (up to 24 months): It is a questionnaire to be filled out by the parent related to communication and other abilities [4].
- Parents' Evaluation of Developmental Status (PEDS) (birth-8 years): It is used for identifying the milestones of the children. It is to be filled by the parents [5]
- Modified Checklist for Autism in Toddlers (MCHAT) (16–30 months): It is a questionnaire to be filled out by the parents and is used to identify children with autism [6].
- Quantitative Checklist for Autism in Toddlers (QCHAT) (18–24 months): It is a questionnaire to be filled out by the parents to identify children with autism [20]
- Screening Tool for Autism in Toddlers and Young Children (STAT) (24–36 months): It is an activity-based assessment to test for various milestones of the children like communication, playing and imitating others [7].

Once the screening is done, if there is any problem in the child's development, they will be directed to diagnostic tools to identify the autism and its severity. Various diagnostic tools available are Autism Diagnosis Interview [8], Autism Diagnostic Observation Schedule [9], Childhood Autism Rating Scale (CARS) [10] and Gilliam Autism Rating Scale [11].

The Literature survey of machine learning algorithms applied to various screening tools is as follows:

1. Kazi Shahrukh Omar and others have developed a mobile application using 2 different datasets. One is an AQ-10 dataset and the other is a real dataset collected from persons. It was able to predict autism with an accuracy of 92.26%, 93.78%, and 97.10% for child, adolescent, and adult respectively for the AQ-10 dataset and 77% to 85% for the real dataset. A combination of random forest CART and random forest ID3 has been used [12].
2. Suman Raj and others have used various machine learning algorithms such as Naïve Bayes, KNN, support vector machine, neural network, logistic regression, and convolutional neural networks on 3 different datasets of ASD screening data for adults, children and adolescents. CNN has the best accuracy on all the datasets compared to other algorithms with an accuracy of 99.53% for adults, 98.30% for children and 96.88% for adolescents [13].
3. Fadi Thabtah and others used logistic regression on QCHAT-10 and AQ-10 datasets of adolescents and adults collected from the ASD test app. Information gain and the Chi-square test for feature analysis have been applied and were able to detect autism and the necessary features to detect autism [14]

4. Fadi Thabtah and others have used the Rules Machine learning algorithm on the AQ-10 dataset which helped to detect autism and also provide various rules which ensure whether the individual is autistic or not [15]
5. Tania and others have applied Adaboost, FDA, C5.0, GLMBOOST, LDA, MDA, PDA, SVM, and CART on QCHAT-10 for toddlers and AQ-10 datasets for adolescents and adults and SVM performed better for the toddler dataset. Before using classifiers various feature transformation techniques such as sine, log, and Z-score have been applied. After classification, feature selection techniques were used which helped to find the important features to detect autism [16]
6. Gennaro and others used various classification algorithms such as random forest (RF), naïve Bayes (NB), support vector machine (SVM), logistic regression (LR), and K-nearest neighbors (KNN) on QCHAT and QCHAT-10 and achieved an accuracy of 95% for SVM [17].
7. Mujeeb Rahman KK and others have used deep neural networks on QCHAT and QCHAT-10 datasets and achieved an accuracy of 97.50% and 100% respectively [18]

The above papers have used various machine learning algorithms to classify the child or adolescent as autistic or not. Fadi Thabtah [15] has used rules-based machine learning to generate an intelligent classifier for autism which is applicable only to adults. In this paper, we propose a rule based classifier model which generates rules used by the screening tool that are combined with machine learning algorithms to detect autism in children of the age group 18–30 months.

3 Proposed Rule Based Classifier Model

The diagram in Fig. 1 shows the architecture of the proposed Rule based Classifier model for detection.

The QCHAT dataset has been collected from the Mendeley data website [19] which is publicly available. The QCHAT dataset is based on the QCHAT screening tool and consists of 25 questions that corresponds to the child's various developmental areas such as speech and language, sensory issues, Eye contact, social communication, lack of

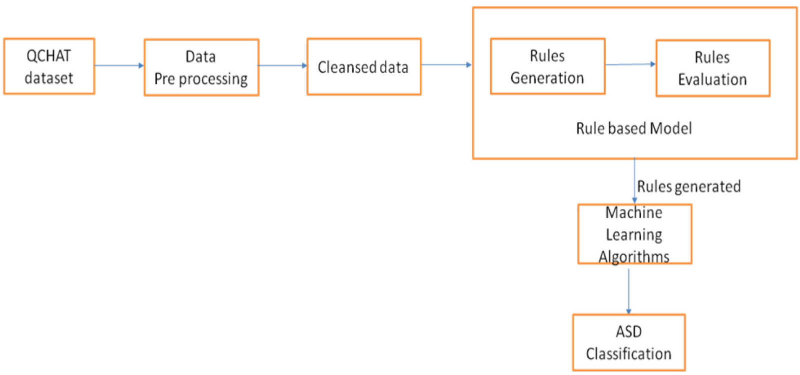

Fig. 1. Rule Based Classifier for the Detection of Autism in Children

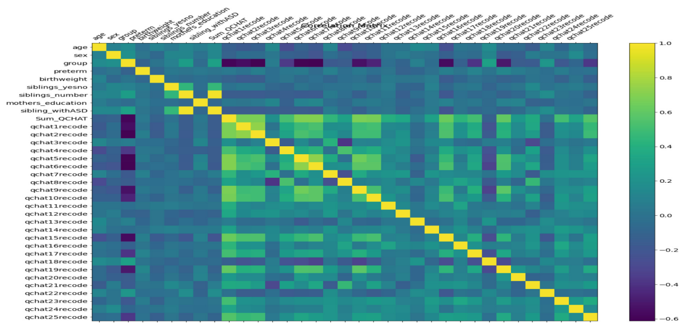

Fig. 2: Correlation matrix of the QCHAT dataset used for feature selection

attention, and concentration. The dataset has 252 records of the children who participated in the screening test with 36 attributes which include personal information of the child such as child_id, age, number of siblings, birth weight, and so on. Once the data is collected, Data Pre Processing is done on the dataset to check for empty values, duplicates or invalid data.

Feature selection is used to find the highly correlated features. The correlation matrix of the dataset is shown in Fig. 2. Based on the correlation matrix we have selected all 25 attributes which represent the questionnaire of QCHAT and other attributes which are highly correlated.

The correlation coefficient of the two variables is obtained by

$$\rho xy = \mathrm{Cov(x, y)}/\sigma x \sigma y \tag{1}$$

where Cov(x,y) is the covariance of the variables x,y

And $\sigma x \sigma y$ are the products of the standard deviation of the variables.

The cleansed data obtained after data preprocessing is sent to rule based model to generate the rules which are then combined with machine learning models for the detection of autism in children. The outcome of the Rule based model is the set of rules generated which when combined with machine learning algorithms classify the new instance as autistic or not autistic.

The algorithm to generate rules is as follows:

Algorithm: Rule Based Classifier for the Detection of Autism in children
Input: QCHAT dataset with n users, min_val,min_strength // Rule_strength=min_strength and minimum frequency=min_value

Output: Set of Rules
1. S_f_R $\leftarrow$ { }
2. r_i $\leftarrow$ { }
3. Temp $\leftarrow$ n
4. DO {
5. If $p(A_i,V_i)\ \forall A_i/\ V_i \geq$ min_val
{6.if $p(A_i,V_i)\ \forall\ i\ /\sum V_i \geq$ min_strength
{7.ri={Ai,Vi}

}
8. Repeat steps 5-7 till no instances of A_i,V_i are found
}
}
9.S_f_R=ri
10. Exit when n has no more instances OR all p(A_i,V_i) instances, have been covered
11. Generate S_f_R
12. Classify Test(Test,S_f_R)

The above algorithm takes QCHAT dataset as an input and the output is the set of rules(S_f_R) generated which when combined with machine learning algorithms can be used to classify ASD or NO ASD. It has n number of users and each user has m number of attributes A_1, A_2, A_3,__,A_m. Each attribute has a value V_i associated with it. The algorithm uses two threshold values min_value and min_strength. The Rule is written as (A_1,V_1)and (A_2,V_2) and (A_3,V_3)and …(A_k,V_k)->C_n where the antecedent is the conjunction of an attribute and its value and the consequent is a class label. Here class labels are ASD or NO ASD. For every user, only if each attribute value is $> =$ min_value, then it can be part of the rule. When the total of all the values is $> =$ min_strength, then it can be added to the set of rules(S_f_R). Initially, S_f_R is empty. As the instances satisfy the conditions, the rules will be generated as shown in steps 5–8.The generated rules will be added to S_f_R in step 9 and this will be repeated for all the instances as shown in step 10. Final S_f_R will be generated in step 11 which are combined with machine learning algorithms to classify the children as ASD or NO ASD.

4 Results

Autism is a Neurodevelopment disorder that affects the child's growth in various aspects such as communication, social skills, eye contact, interpersonal skills, language, and sensory issues. In the proposed system autism is detected using Rule based classifier. QCHAT dataset which is publicly available is used for this purpose. It consists of 252 records with 36 columns. The metrics used for detection are accuracy, precision, and recall.

Accuracy is a measure of how well a machine learning algorithm can make correct predictions.

$$Accuracy = Number\ of\ correct\ predictions / Total\ number\ of\ predictions \tag{2}$$

Precision is a measure of the number of true positives to the total number of predictions.

$$Precision = True\ Positives / (True\ positives + False\ Positives) \tag{3}$$

Recall is a measure of number of the true positives to the total number of positive predictions.

$$Recall = True\ Positives / (True\ positives\ False\ Negatives) \tag{4}$$

Table 1. Confusion matrix and ROC curve for Random forest classifier

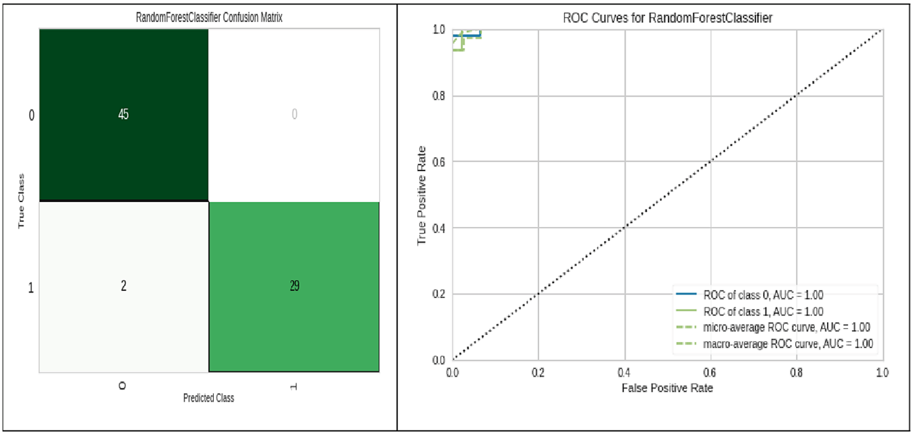

Table 2. Metrics of Random forest on train and test dataset

Data	Accuracy	Precision	Recall
Training Data	96.6%	98.5%	93.2%
Testing Data	97.37%	100%	93.5%

The model was developed by using Google Colaboratory provided by Google which facilitates research using machine learning. It has a Jupiter notebook which is cloud-based and can be accessed without any setup needed. Upon generation of rules, a Random forest algorithm has been applied to it. The dataset is divided into training and testing with a 70:30 ratio respectively and 10-fold cross-validation is used. The confusion matrix and the ROC curve are shown in Table 1.

The output of random forest classifier is a dataset that consists of only autistic children which are around 76 rows and 37 columns where rows indicates the number of children who are autistic and columns indicate the attributes of the QCHAT dataset with an additional attribute of autistic. The autistic attribute has a value of 1 if the child is autistic and a 0 if not autistic.

The Accuracy, Precision, and Recall for the training data and testing data are as shown in Table 2.

5 Conclusion

Autism is a developmental disorder that affects the child's growth in various aspects and early detection helps to develop the necessary skills. Early detection can help the child to develop to their fullest. Evaluation of screening tools is carried out manually by the doctors. Using machine learning approaches in the detection helps doctors to detect autism in less time and with better accuracy avoiding human errors. Our proposed system of Rule based classifier is the first of its kind to combine rules which are used by the screening tool and machine learning algorithms to detect autism effectively with an accuracy of 97.37%.

One of the limitations is the size of the dataset. The proposed method can achieve higher accuracy if more data can be collected. Our future scope is to combine machine learning algorithms for diagnostic tools of autism to increase their performance and reduce evaluation time.

References

1. Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 years of age. Arch Gen Psychiatry. 2006 Jun;63(6):694–701. doi: https://doi.org/10.1001/archpsyc.63.6.694. PMID: 16754843.
2. CDC website, https://www.cdc.gov/ncbddd/autism/index.html
3. ASQ website, https://agesandstages.com/screening-navigator/screening
4. Brookes Website, https://brookespublishing.com/product/csbs-dp-itc/
5. Pedstest Website, http://www.pedstest.com/
6. Mchatscreen website, https://mchatscreen.com/
7. Vanderbilt website, http://vkc.mc.vanderbilt.edu/vkc/triad/training/stat/
8. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994 Oct;24(5):659–85. doi: https://doi.org/10.1007/BF02172145. PMID: 7814313.
9. Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000 Jun;30(3):205–23. PMID: 11055457.
10. Chlebowski, C., Green, J.A., Barton, M.L. *et al.* Using the Childhood Autism Rating Scale to Diagnose Autism Spectrum Disorders. *J Autism Dev Disord* 40, 787–799 (2010). https://doi.org/10.1007/s10803-009-0926-x
11. Samadi SA, Noori H, Abdullah A, Ahmed L, Abdalla B, Biçak CA, McConkey R. The Psychometric Properties of the Gilliam Autism Rating Scale (GARS-3) with Kurdish Samples of Children with Developmental Disabilities. Children (Basel). 2022 Mar 19;9(3):434. doi: https://doi.org/10.3390/children9030434. PMID: 35327806; PMCID: PMC8947096.
12. K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi and M. N. Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder," *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 1–6, doi: https://doi.org/10.1109/ECACE.2019.8679454.
13. Suman Raj, Sarfaraz Masood, Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, Procedia Computer Science, Volume 167,2020,Pages 994–1004, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.03.399.
14. 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 2017 (ICMHI '17). Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3107514.3107515
15. Thabtah, F., & Peebles, D. (2020). A new machine learning model based on induction of rules for autism detection. *Health Informatics Journal*, 264–286. https://doi.org/10.1177/1460458218824711

16. T. Akter *et al.*, "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," in *IEEE Access*, vol. 7, pp. 166509–166527, 2019, doi: https://doi.org/10.1109/ACCESS.2019.2952609
17. Tartarisco G, Cicceri G, Di Pietro D, Leonardi E, Aiello S, Marino F, Chiarotti F, Gagliano A, Arduino GM, Apicella F, Muratori F, Bruneo D, Allison C, Cohen SB, Vagni D, Pioggia G, Ruta L. Use of Machine Learning to Investigate the Quantitative Checklist for Autism in Toddlers (Q-CHAT) towards Early Autism Screening. Diagnostics (Basel). 2021 Mar 22;11(3):574. doi: https://doi.org/10.3390/diagnostics11030574. PMID: 33810146; PMCID: PMC8004748
18. Mujeeb Rahman KK, Monica Subashini M. A Deep Neural Network-Based Model for Screening Autism Spectrum Disorder Using the Quantitative Checklist for Autism in Toddlers (QCHAT). J Autism Dev Disord. 2022 Jun;52(6):2732-2746. doi: https://doi.org/10.1007/s10803-021-05141-2. Epub 2021 Jun 30. PMID: 34191261.
19. Niedźwiecka, Alicja; Pisula, Ewa; Domasiewicz, Zuzanna (2019), "Q-CHAT scores of Polish toddlers with autism spectrum disorders and typically developing controls", Mendeley Data, V1, doi:https://doi.org/10.17632/tmpkt2mfkg.1(QCHAT dataset).
20. Allison C, Baron-Cohen S, Wheelwright S, Charman T, Richler J, Pasco G, Brayne C. The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): a normally distributed quantitative measure of autistic traits at 18–24 months of age: preliminary report. J Autism Dev