ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/361373188

Diagnosing autism in low-income countries: Clinical record-based analysis in Sri Lanka

Article *in* Autism Research · June 2022

DOI: 10.1002/aur.2765

CITATIONS
0

READS
16

7 authors, including:

Chitraka Wickramarachchi
University of Sri Jayewardenepura
7 PUBLICATIONS **69** CITATIONS
SEE PROFILE

Pradeepa Samarasinghe
Sri Lanka Institute of Information Technology
39 PUBLICATIONS **77** CITATIONS
SEE PROFILE

D. M. A. Dahanayake
University of Colombo
28 PUBLICATIONS **53** CITATIONS
SEE PROFILE

Madhuka Nadeeshani
Sri Lanka Institute of Information Technology
5 PUBLICATIONS **10** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Final year project-Data Science View project

professor nalin da sival View project

All content following this page was uploaded by Madhuka Nadeeshani on 20 June 2022.

The user has requested enhancement of the downloaded file.

Received: 6 June 2021 | Accepted: 20 May 2022
DOI: 10.1002/aur.2765

RESEARCH ARTICLE

Diagnosing autism in low-income countries: Clinical record-based analysis in Sri Lanka

Hashan Peiris[1] | Darshana Chitraka Wickramarachchi[1] | Pradeepa Samarasinghe[2] | Philip Vance[3] | Dulangi Dahanayake[4] | Veerandi Kulasekara[2] | Madhuka Nadeeshani[2]

[1]Department of Statistics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
[2]Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
[3]Intelligent Systems Research Centre, University of Ulster, Londonderry, UK
[4]Faculty of Medicine, University of Colombo, Colombo, Sri Lanka

Correspondence
Pradeepa Samarasinghe, Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.
Email: pradeepa.s@sliit.lk

Funding information
World Bank

Abstract

Use of autism diagnosing standards in low-income countries (LICs) are restricted due to the high price and unavailability of trained health professionals. Furthermore, these standards are heavily skewed towards developed countries and LICs are underrepresented. Due to such constraints, many LICs use their own ways of assessing autism. This is the first retrospective study to analyze such local practices in Sri Lanka. The study was conducted at Ward 19B of Lady Ridgeway Hospital (LRH) using the clinical forms filled for diagnosing ASD. In this study, 356 records were analyzed, from which 79.5% were boys and the median age was 33 months. For each child, the clinical form together with the Childhood Autism Rating Scale (CARS) value were recorded. In this study, a Clinically Derived Autism Score (CDAS) is obtained from the clinical forms. Scatter plot and Pearson product moment correlation coefficient were used to benchmark CDAS with CARS, and it was found CDAS to be positively and moderately correlated with CARS. In identifying the significant variables, a logistic regression model was built based on clinically observed data and it evidenced that "Eye Contact," "Interaction with Others," "Pointing," "Flapping of Hands," "Request for Needs," "Rotate Wheels," and "Line up Things" variables as the most significant variables in diagnosing autism. Based on these significant predictors, the classification tree was built. The pruned tree depicts a set of rules, which could be used in similar clinical environments to screen for autism.

Lay Summary

Screening and diagnosing autism in low-income countries such as Sri Lanka has always been a challenge due to limited resources and not being able to afford global standards. Due to these challenges, locally developed clinical forms have been used. This study is the first to analyze a clinical record set for autism in Sri Lanka to benchmark the local clinic form with a global standard. Furthermore, this study identifies the most significant diagnostic symptoms for children and based on these significant features, a simple set of IF–THEN rules are derived which could be used for screening autism in a similar clinical environment by health officials in the absence of consultants.

KEYWORDS
ASD, ASD diagnosing standards, ASD predictors, autism, CARS, classification, cultural factors, logistic regression, low-income countries

© 2022 International Society for Autism Research and Wiley Periodicals LLC.

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that causes significant challenges in social communication and social interaction, along with repetitive and restrictive behaviors (Birx et al., 2011). As the symptoms vary from mild to severe, ASD is known as a spectrum disorder. According to the Centers for Disease Control and Prevention (CDC)'s Autism and Developmental Disabilities Monitoring (ADDM) Network statistics, 1 in 54 children were reported to have autism in the United States, 2016 (Maenner et al., 2020). Worldwide epidemiological surveys emphasize that there is a rising prevalence in diagnosis of ASD over recent years (Fombonne, 2018; Maenner et al., 2020).

Sri Lanka (SL) is a low-income country (LIC) in the Southern Asia region. South Asia represents more than 20% of the world's population, yet prevalence studies on autism have been limited and not conducted consistently. During the years 2011–2012, a study has been conducted in five districts across India to identify different neurodevelopmental disorders among 2- to 9-year-old children (Arora et al., 2018). While the five sites have varying prevalence rates, it was found that the overall prevalence rate of autism is comparable with the global prevalence. Although a prevalence study has not been conducted in SL, a small community survey was undertaken in 2007; the results showed that the prevalence rate for autism is 1 out of 93 children (Perera et al., 2009). Furthermore, it was found that by the age of 24 months, only 14.3% of children with ASD in SL had been diagnosed (Perera et al., 2013). Consequently, this delayed identification has had a significant impact on the access to early intervention and adversely affects developmental outcomes (Perera et al., 2013).

Though research on autism has advanced in the last decade, this was concentrated in high income countries, overlooking the LICs as well as the ethnic minorities (de Leeuw et al., 2020). A comparative study conducted in Japan, India, and the UK, identified the importance of developing screening tools for ASD in cross-cultural context. (Carruthers et al., 2018). This requirement is elaborated in de Leeuw et al. (2020), developing a model to account for cultural and contextual factors at all levels of ASD. The framework proposed in de Leeuw et al. (2020) discusses that although broad symptoms of autism are universally observed, cultural differences have been highlighted in the manifestation of autism qualitatively as well as quantitatively. Furthermore, the analysis points out that the structure of ASD diagnosis tools, Diagnostic and Statistical Manual fifth edition (DSM-5) (American Psychiatric Association, 2013) and International Classification of Diseases (ICD-11) (World Health Organization [WHO], 2020), were supported by the factor analysis conducted in high-income western countries and little is known whether the cross-cultural effects have been accounted in the current diagnostic standards.

No in-depth study has been conducted on the suitability of the global standards to the cultural and social differences in SL, but the study in Perera et al. (2009) used a Modified Checklist for Autism in Toddlers (MCHAT) test, which was translated to Sinhala language for screening. In this study, as it was found that MCHAT was only 25% sensitive, it was recommended to develop assessment tools based on socio-cultural norms. In addressing this problem, a paper-based screening tool incorporating written and pictorial content has been developed and is referred to as Pictorial Autism Assessment Schedule (PAAS) (Perera et al., 2017). PAAS prediction is based on the parent's responses to a set of questions. So far, no predictive tool has been developed in SL for screening ASD based on the clinical assessments used in local setting. Specifically, the predictive tool is a statistical model, which is used to forecast the status (having high risk for ASD or low risk for ASD) of a child using the child's social, communication, and cultural indicators.

According to the conceptual framework suggested in de Leeuw et al. (2020), subtle differences due to cultural and contextual setting can be seen in levels of expression, recognition, interpretation, and reporting of autism symptoms. The cultural beliefs such as autism being "the western disease" and supernatural explanations given in de Leeuw et al. (2020) are common in SL too. As an example, the interpretation for speech delay would be due to "inauspicious timing of the first hair cut." Further, beliefs such as "boys usually start speaking late" might be due to lack of awareness on children reaching their growth milestones (Bhavnani et al., 2021).

Though the government of SL assumes most of the health service cost at present, there is a scarcity of state-sponsored health programs in the community for ASD. A few private sector facilities are available, but for most families these remain inaccessible because of the high cost. Additionally, ASD diagnosis requires multi-disciplinary resources, which are not available in many parts of the country (Perera et al., 2016). The SL Health Ministry statistics in 2016 (Ministry of Health Nutrition and Indigenous Medicine, 2016) highlighted that many districts have neither child psychiatrists nor developmental pediatricians to diagnose ASD at early stages. In 2011, according to Chandradasa and Kuruppuarachchi (2017), there were only 0.29 psychiatrists per 100,000 population within SL. These factors corroborate the claims made in the fourth level of framework in de Leeuw et al. (2020), as affordability and availability are two key factors directly affecting the screening rate of ASD within the country.

Lady Ridgeway Hospital for Children (LRH) is currently the largest free-of-charge Pediatric Hospital in SL (Lady Ridgeway Hospital [LRH], 2020). At present, child and adolescent psychiatrists and trained medical officers at LRH assess children for ASD and record their findings using paper-based forms, which have evolved over time. The records comprise of raw data concerning local factors and diagnoses, which have remained largely untouched. To this end, an extensive analysis in this study was carried out on the clinical records gathered at child and adolescent mental health services at LRH, to assess how it could be extended to the whole country as a predictive ASD screening tool.

The overall objectives of this study were:

1. To validate the clinical form used at LRH based on clinical records, taking Childhood Autism Rating Scale (CARS) as the benchmark.
2. To develop a model for identifying the key predictors of ASD in young children of SL, using a dynamic clinical dataset. It is a key objective that this model will use the most appropriate and reliable symptoms, taking account of cultural, social, and behavioral patterns.
3. To build a predictive rule-based model for screening of children for ASD, which could be used in any hospital with minimal intervention of consultants.

A logistic regression model is built to achieve the second objective: to find the significant factors associated with ASD. The Logistic regression model contains many predictors including interaction terms. In order to predict the status of a child, one has to compute the probability of a child having high risk for autism using the complex model, which is not easy, especially in rural clinics where there are no computer facilities. Therefore, a classification tree, a simple IF–THEN rule set, is built under the third objective using the significant variables identified in the logistic regression model to predict the status of a child. More specifically, in classification trees, a series of conditions are checked on predictors (IF statements) before predicting the class (Then statement). Reader may refer to the section Results: Classification Tree for an example of IF–THEN rule set and more details of classification trees are given in the section Methods: Classification Tree.

METHODS

Context

The diagnostic manuals such as DSM-5, ICD-10 (World Health Organization [WHO], 1992), and assessment tools such as CARS (Chlebowski et al., 2010) and Autism Diagnostic Observation Schedule (ADOS) (Western Psychological Services, 2012) are the most popular criteria-based evaluations used by the clinicians to diagnose ASD (Petrocchi et al., 2020; Thabtah & Peebles, 2019). The affordability of purchasing such standard assessment tools to cater to the general public is minimal in a LIC like Sri Lanka. A comparative review is carried out on DSM-5, ICD-10, and CARS to understand the feasibility of applying these tools in assessing children with ASD in Sri Lanka and is attached as a supplementary document. As detailed in the review, due to limited expertise and validation studies conducted for cross cultural contexts in ICD-10 and DSM-5, applying those effectively across the population is questionable.

Participants and materials

This is a retrospective study conducted at Ward 19B of LRH using the completed clinical forms for diagnosing ASD. During the child's first visit to the clinic, the parent/caregiver will be interviewed by the medical officers and the behavior of the child is observed. The responses and observations are recorded in the clinical forms, which include sections for personal demographics, medical history, behavioral problems, social observations, language and communication observations, and repetitive and stereotyped behavior observations. The CARS is also used to rate behavior. The completed clinical form and the CARS sheet will then be evaluated by the consultant following the same process: interviewing the caregiver and observing the child where the consultant may note any added/contradicting points resulting in the final ASD diagnosis status. The diagnostic assessment for each child was carried out according to the DSM-IV criteria.

Inclusion/exclusion criteria

In this study, the CARS sheets and the clinical forms of the children who visited Ward 19B for their first assessment of ASD from October 2014 to March 2021 were reviewed. All the children who had a complete clinical form and a CARS sheet at the first diagnosis assessment have been considered under this analysis. Altogether 356 entries were recorded, however, there were missing values in the dataset. For some records the missing values were imputed and the others which have a higher number of missing values were removed from the dataset.

The ethics approval from Ethics Review Committee of LRH was obtained to review the data that had already been collected as part of the clinical service. As this is a retrospective study and no further contact was made with the diagnosed children or the caregivers, the informed consent of the individual participants were waived. Thus, according to the ethics approval, we maintain the anonymity of the data throughout the paper.

Procedure

In Ward 19B, the data of the participants have been recorded within hand-filled forms and stored in separate folders for each participant. Initially the folders were examined and the folders having both LRH form, and the CARS sheet were separated. Then an Excel spreadsheet was created to capture the data from the medical forms. In the dataset, data was recorded regarding the child's clinical number, date of birth, sex, age at first assessment, medical history, speech development, schooling stage and other behavioral issues such as sleeping disorders, hyperactivity, and temper tantrums. Other than these, the observations made by the health officials were also recorded. Finally, each record included CARS assessment as well. The different sections of the LRH form together with the recorded parameters are detailed in supplementary document. Two research assistants were involved in recording the data. Further, a quality control was conducted through a second review by a different group.

Measures

Demographics

Demographic information of the children under study were extracted from the clinical forms such as clinical number, date of birth and sex. The age of the participants varied between 12 and 68 months having a median age of 33 months. In the initial data set of 356 records, 79.5% records were males. However, after removing observations with missing values, the data set was reduced to 317, of which approximately 78.5% were boys and the minimum age was 14 (Table 1).

Clinical measures

Under the medical history of the participants, diagnoses of epilepsy and gastrointestinal symptoms were recorded. Out of the 356 participants, 15 had suffered from epilepsy while 10 had gastrointestinal symptoms. Since this study is based mainly on the features recorded under observation, three types of observations were considered for building the models: Social Relatedness, Use of Language and Communication, and Repetitive and Stereotyped Behavior. Seventeen diagnostic variables under these three types together with age at first assessment and gender were considered as predictors, while the CARS was considered the response variable.

The predictors under Social Relatedness: "Eye contact," "Response to name," "Interaction with parents," "Interaction with others" and "Pointing," were quantitative variables. While "Request for needs," "Talking to self," "Echolalia," "Echopraxia" under the Use of Language and Communication category were dichotomous variables; "Response to commands" was a polytomous variable. All the predictors in Repetitive and Stereotyped Behavior category: "Flapping of hands," "Rotate wheels," "Looking at fingers," "Line up things," "Interest in specific toys," and "Spinning" were dichotomous variables. In addition to these, the variables "Age at first assessment" was a quantitative variable while "Gender" and "Speech development" were dichotomous variables.

Statistical analysis

In analyzing the predictors, it was found that some of the entries in forms had not been recorded, resulting in a number of missing values. Hence, imputation was carried out to replace the missing values. Once all the missing values were imputed, the consistency of CARS and Clinically Derived Autism Score (CDAS) were evaluated. Logistic regression was then used to find the clinically significant variables. Finally, based on the significant variables identified, a rule set was constructed to screen children with ASD.

Missing value imputation

The clinical forms were filled by different medical officers and depending on the expertise and experience there could be personal differences in filling the data. Furthermore, the initial dataset of 356 observations had missing values across different variables. For example, it was observed that all the predictors under the Repetitive and Stereotyped Behavior category had between approximately 1% and 5% of missing values. This may have been due to medical officers being unable to observe the stereotypical behavior during the clinical observations, hence being hesitant in providing a decision on the indicator.

Too many imputations of missing values for either an observation or a predictor is not recommended (Scheffer, 2002); Allison (2001) recommends at most 25% of imputations for an observation and at most 10% of

TABLE 1 Age and CARS score distribution statistics

Parameter	Min	Q-1	Median	Q-3	Max	Std. dev
Age at first assessment	14	25.0	33.0	38	68.0	9.50
CARS						
a. Low risk	17	23.5	26.0	28	29.5	3.44
b. High risk	30	34.0	37.25	41	56.0	5.02

Abbreviation: CARS, Childhood Autism Rating Scale.

TABLE 2 Arguments of the model for data type in "mice()"

Argument value	Variable type
PMM	Numeric variables
logreg	Factor variables with two levels
polyreg	Factor variables with more than two levels
polr	Ordinal variables with more than two levels

imputations for a predictor so as not to adversely affect the final results. Hence, an observation was removed if it contained missing values for more than 25% of predictors, resulting in 28 observations being removed. A predictor was deleted if it contained missing values for more than 10% of observations resulting in the predictor variable, "Echopraxia," being removed.

Multivariate Imputation by Chained Equations (MICE) algorithm was used to impute the multivariate missing data. In the MICE algorithm, each predictor which contained missing values (Incomplete Predictor) was linearly regressed by remaining predictors in the data set. Essentially, the incomplete predictor became the response variable in the regression model. The regression methods used were determined by the variable type of the incomplete predictor; this study contained a combination of numerical/dichotomous and polytomous variables. The "MICE package" in R was used to impute missing values in the predictors. The regression methods available in the MICE package are shown in Table 2.

Assessing the consistency between CARS and the CDAS

The recorded children for this study were assessed with both the CARS standard and clinical observation based on predictors independently. Hence the first objective was to test the consistency between the CARS and CDAS, where the latter is based on clinically observed predictors.

The standard CARS form contains 15 ordinal questions. Each question is answered in 1–4 scale, where 1 and 4 represent the least and most severe levels of the characteristic being observed, respectively. On the other hand, some of the clinically observed predictors are quantitative with values ranging from 0 to 100, and others are qualitative with dichotomous and polytomous levels. To calculate the total score for the clinical form, the qualitative variables needed to be transformed to "numerical sensed" variables. The numerical values were selected between 1 and 4 in order to be consistent with the CARS form.

The "Response to commands" predictor was assigned numerical values according to the natural ordering of its levels.

$$\text{Response\_to\_commands}_{\text{new}} = \begin{cases} 1 \text{ if } X = \text{Always} \\ 2 \text{ if } X = \text{Often} \\ 3 \text{ if } X = \text{Seldom} \\ 4 \text{ if } X = \text{Never} \end{cases} \tag{1}$$

Next, to align with the range of CARS, dichotomous predictors were transformed according to Equation (2) while a linear mapping was performed on each quantitative variable as in Equation (3).

For dichotomous predictors, the range 1–4 was divided to three equally spaced intervals as 1–2, 2–3, and 3–4. The "NO" category was assigned the value 1.5 (mid value of the first interval) and "YES" category was assigned 3.5 (mid value of the third interval) (Equation 2). As the quantitative variables in the clinical form have values ranging from 0 to 100, a linear transformation is applied to make the range between 1 and 4 (Equation 3). Hence, these transformations ensured the new predictor values were within 1–4 range.

1. If X was a dichotomous type variable

$$X_{\text{new}} = \begin{cases} 1.5 \text{ if } X = \text{'No'} \\ 3.5 \text{ if } X = \text{'Yes'} \end{cases} \tag{2}$$

2. If X was a quantitative type variable

$$X_{\text{new}} = 1 + \frac{3}{100}X \tag{3}$$

However, it is important to state that the model building (second and third objectives) was done on the original variables.

Once all clinically observed predictors were transformed, the sum of the values is taken as the measure of the presence and severity of ASD. Scatter plot and Pearson product–moment correlation coefficient are used to assess the linear relationship between two types of scores (CARS and CDAS), which evaluate the consistency of the two scores.

Identifying clinically significant variables: Logistic regression

When a child presents to the LRH clinic, he/she is evaluated by the observations made by the medical officer based on indicators under the categories Social Relatedness, Use of Language and Communication, and Repetitive and Stereotyped Behavior. Since we are interested in identifying clinically significant variables on the presence of ASD, a logistic regression model was fitted to identify the clinically significant factors, which could be used as determinants for diagnosing ASD in LRH Clinic. Dichotomized CARS score (Equation 4) was used as the response variable (Y).

$$Y = \begin{cases} 0(\text{Low Risk})\text{ if CARS} \le 30 \\ 1(\text{High Risk})\text{ if CARS} > 30 \end{cases} \tag{4}$$

According to DSM-5, "Echolalia" is categorized under the repetitive or stereotyped development, and it is found that the repetitive behaviors are observed later in comparison to social interactions and communication. According to Roberts (1989), echolalia is a typical behavioral characteristic in a child below 24 months, however, within 30–36 months, presence of echolalia is associated with language delay. Hence "Echolalia" cannot be reliably measured of a child who is less than 30 months of age.

Accordingly, "Echolalia" should be recorded as "NO" for a child who is less than 30 months of age. However, it was observed that in the records "Echolalia" contains both "YES" and "NO" responses for children who are less than 30 months of age. Therefore, "Echolalia" was purposely changed to "NO" for children younger than 30 months that had already been recorded as "YES." To distinguish artificially changed "NO" and genuinely recorded "NO" (for those who are older than 30 months) an indicator variable, "Indicator," was created by recording all observations with "Age at first assessment" less than or equal 30 months as 1 and 0 otherwise.

Response to command was measured in ordinal scale at the clinic. However, when including this predictor into the logistic regression model it was dichotomized (Equation 5).

$$\text{Response to command} = \begin{cases} 0\text{ if Never and Seldom} \\ 1\text{ if Always and Often} \end{cases} \tag{5}$$

Dichotomized CARS was modeled using a logistic regression model which is given in Equation (6). The exponent of the coefficient of jth variable (Equation 7) gives the odds ratio of being in the high-risk category for a unit increment of the jth predictor while other predictors are held constant.

$$\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^{q} \beta_i X_i \tag{6}$$

$$\frac{\left(\frac{P|X_j=x_j}{1-P|X_j=x_j}\right)}{\left(\frac{P|X_j=x_j+1}{1-P|X_j=x_j+1}\right)} = \exp(\beta_j) \tag{7}$$

where, p is the probability of a child with ASD given his/her predictor X_j when other predictors held constant while q represents the number of predictors in the model.

The distribution of age at first assessment and the CARS score for the data set are given in Table 1.

Akaike Information Criterion (AIC), null deviance and residual deviance were used to select the best fitted model (Pace & Briggs, 2009). A model with lower values for these three statistics indicates a better fit for the data.

Classification tree

The third objective of the study is to derive rules to classify children who are suspected of having ASD; this could be achieved by constructing a classification tree. In a binary classification tree, the predictor space is recursively partitioned (or split) into two disjoint sub-regions (or nodes) until each sub-region becomes homogeneous (terminal nodes) with respect to a class, either Low Risk or High Risk. At each split, the impurities of two resultant sub-regions are calculated for each predictor variable using an impurity function. The impurity reduction is then calculated as the difference between the node impurity and the sum of the weighted impurities of the resultant nodes. The split which gives the largest impurity reduction is then taken as the best split for the node. In this research, the GINI impurity function was used as it is commonly used in medical research (Shouman et al., 2011). Out of the many types of classification trees available, such as CART and C4.5; CART is used in this study.

Generally, a fully grown tree overfits the data (Frank, 2000); to get a better fit, unnecessary branches are pruned. There are two types of pruning namely: (1) Post-pruning and (2) Pre- pruning. In post-pruning, the pruning is done after the full tree is built. Pre-pruning is done while the tree is being built. This study used the pre-pruning to create the best pruned subtree, which was then used to derive rules to screen ASD cases.

Model evaluation

Model evaluation is an integral part of statistical modeling, as it provides insight into the extent the model generalizes the population being studied. Out of many model evaluation methods (Bramer, 2007), the independent test set approach is the most recommended method, provided that a sufficiently large test set exists (Martens & Dardenne, 1998). In this study, the observed sample is partitioned to 70:30, where 70% of observations was used to fit the model, and the remaining 30% was used to test the model. This process was repeated 1000 times so that model evaluation measures could be obtained with the respective standard errors. The misclassification rate, sensitivity (true positive rate) and specificity (true negative rate) were used as model evaluation measures. This procedure was applied to evaluate both the models: logistic regression model and the classification tree model.

RESULTS

Missing value imputation

Imputation returned a data set with 328 complete observations on 18 predictors. The CARS variable (the response) also contained missing values, resulting in the deletion of a further 11 observations from the dataset. This resulted in a dataset of 317 observations, where 268 observations were classified as "High Risk."

CARS and CDAS

The scatter plot of CARS versus CDAS in Figure 1 shows a positive linear relationship between the two variables. However, the Pearson product–moment correlation coefficient of 0.45 meant the strength of the relationship was moderate. Hence, Logistic Regression, instead of a multiple linear regression approach was used in finding the clinically significant predictors.

Logistic regression

A logistic regression model was fitted on the original predictors to find the significant factors which could help predict ASD. As there was no predetermined set of predictors, the fullest possible model was fitted for the data. In addition to the main effects, we tried to include as many intersections as possible to the model. However, due to the limited number of observations, interaction effects involved with more than three factors were not considered. Even some three-way interactions were not estimated due to sparseness. The fitted model showed there are only two two-way interactions which are significant, namely: "Speech: Line up things" and "Rotate Wheels: Line up things." Further, "Response to command: Interest in specific toys" interaction was marginally significant.

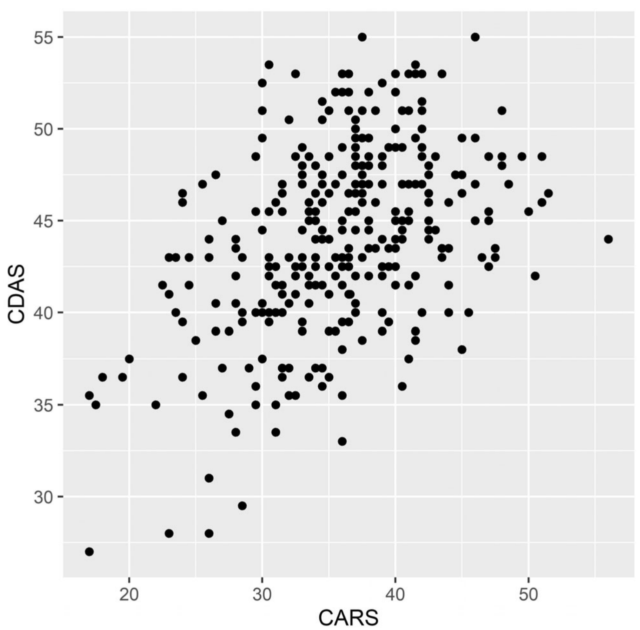

FIGURE 1 Scatter plot of CARS versus CDAS. CARS, Childhood Autism Rating Scale; CDAS, Clinically Derived Autism Score

Parameter estimates, the standard errors, and their significance at the 5% level are given in Table 3. Apart from the significant predictors in the model, "Response to name" is significant at the 10% level of significance. Summary measures for the regression model are given in Table 4.

TABLE 3 Summary of the model

	Estimate	Std. error
Intercept	3.739*	1.743
Age at first assessment	0.003	0.040
Gender—Male	0.631	0.503
Eye contact	−0.032*	0.010
Response to name	−0.019	0.010
Interaction with parents	−0.013	0.009
Interaction with others	−0.034*	0.012
Pointing	0.012	0.007
Speech development-normal	1.325	0.022
Response to commands—YES	0.344	0.627
Looking at fingers—YES	−0.869	0.576
Flapping of hands—YES	1.496*	0.534
N Echolalia—YES	−0.008	0.633
Indicator	−1.076	0.755
Request for needs—YES	−1.399*	0.488
Talking to self—YES	−0.084	0.462
Rotate wheels—YES	1.420*	0.653
Line up things—YES	2.611*	0.855
Interest in specific toys—YES	1.120	0.581
Spinning—YES	0.018	0.453
Response to command—YES: Interest in specific toys—YES	−1.647	0.881
Rotate wheels—YES: Line up things—YES	−2.879*	1.037
Speech normal: Line up things—YES	−3.912*	1.436

*Indicates significant variables at 5% level of significance.

TABLE 4 Model evaluation measures of the logistic regression model

Measure	Estimated value	95% confidence interval
Accuracy	0.85	[0.79, 0.91]
Sensitivity	0.93	[0.86, 0.99]
Specificity	0.38	[0.15, 0.62]

According to the estimated coefficients of the model, the odds of being determined a high risk for autism is increased by 50% for children who enact "Flapping hands" and 40% for children who do not "Request for needs." An increase in "Eye contact" and "Interaction with others," however, decrease the odds of being determined high risk for autism.

In conclusion, the "Eye contact," "Interaction with others," "Flapping hands," "Request for needs," "Rotate wheels" and "Line up things" are the predictors that the study points to as significant in diagnosing ASD.

Classification tree

The classification tree was built on the significant predictors (at 5% and 10% levels of significance) of the Logistic Regression model. These significant terms are given in Table 3.

The tree is grown with pre-pruning and the best-pruned subtree is shown in Figure 2. The rules extracted from the classification tree are listed below.

Rule set for screening low-risk ASD

- A child with "Eye contact" greater than or equal to 38 and "Interaction with others" greater than or equal to 53.

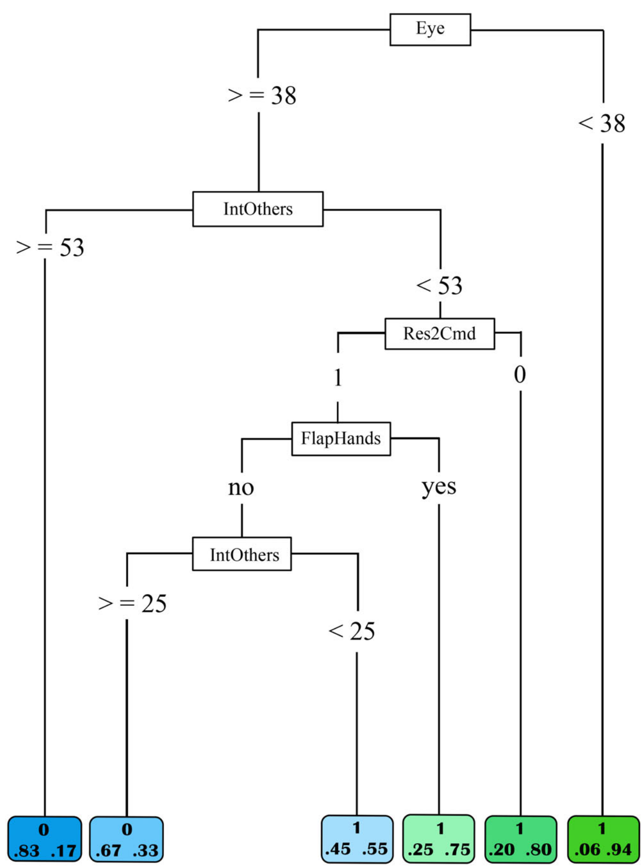

FIGURE 2 Classification tree

- A child with "Eye contact" greater than or equal to 38, "Interaction with others" greater than or equal to 25, who "Responds to commands" and no "Flapping hands."

Rule set for screening high-risk ASD

- A child with "Eye contact" less than 38.
- A child with "Eye contact" greater than or equal to 38, "Interaction with others" less than 53 and no "Response to commands."
- A child with "Eye contact" greater than or equal to 38, "Interaction with others" less than 53, who "Responds to commands" and with "Flapping hands."
  A child with "Eye contact" greater than or equal to 38, "Interaction with others" less than or equal to 25, who "Responds to commands" and no "Flapping hands."

Tree model evaluation measures are given in Table 5.

In summary, a logistic regression model was built to identify the significant factors associated with ASD, containing 23 predictors including three interaction terms. In certain scenarios, such as predicting the status of a child in rural clinics, the use of this model may be difficult as it involves computation. Thus, an alternative classification model is also presented which is a simple predictive tool that can be more easily utilized under these scenarios without the burden of computation.

DISCUSSION

When a child has been referred to LRH for suspected ASD diagnosis, psychiatrists and trained medical officers assess the child and complete paper-based forms. The study revealed that a score computed from a locally developed assessment has a positive and moderate linear relationship with CARS.

In identifying the most valid and reliable symptoms for the accurate diagnosis of ASD, a logistic regression model was fitted by dichotomizing the CARS values. The analysis shows that "Eye contact," "Interaction with others," "Flapping of hands," "Request for needs," "Rotate wheels," and "Line up things" were significant variables in diagnosing ASD. It was also found that the presence of interaction effects of "Rotate wheels" and "Line up things," delays in "Speech" and "Line up things"

TABLE 5 Model evaluation measures of the tree model

Measure	Estimated value	95% confidence interval
Accuracy	0.85	[0.79, 0.91]
Sensitivity	0.94	[0.87, 1.00]
Specificity	0.34	[0.09, 0.59]

and "Response to commands" together with "Interest in specific toys" affect the probability of having ASD. Identified significant predictors were then used to construct a set of classification rules to screen a child for ASD, which could be used in any hospital with minimal intervention of health officials.

In terms of the significant variables, "Response to name" which is considered as one of the key indicators for ASD in global standards has not been identified as a primary feature, however it is marginally significant. A special note on this indicator in the local setting is that most of the time the children are not addressed by their name, but by their pet names. In clinical assessments, sometimes these pet names were also used but there is not enough evidence to distinguish between these in this study. Further, in SL, speech delay has been identified as the most common presenting symptom for ASD (Dahanayake et al., 2015; Perera et al., 2013). However, this study does not reveal "Speech Delay" as a salient feature in diagnosing ASD. This may be because parents are not experienced or aware about the behavioral differences of other variables such as eye contact, which do show significant association with ASD. Moreover, though there is a belief that due to collectivist culture, children in SL are shy and might show subdued behavior in eye contact, in this study it is revealed that "Eye contact" is one of the most prominent symptoms.

Applying these findings to screen children for ASD can be done in two stages. At the first stage, the classification rules will be used for screening children at clinics. Simplicity being an important aspect of the classification tree, a simple printed version of the tree (a tree type diagram) can be used in screening the children at the clinics. Through this printed tree diagram, the first level screening can be carried out by non-specialists, where the children with high risk for ASD will be identified. These high-risk children will then be directed to the specialists for further diagnosis.

As there is only a limited number of child and adolescent psychiatrists in SL (11 as at 2021) (Postgraduate Institute of Medicine [PGIM] Board Certified Specialists – Child & Adolescent Psychiatry, n.d.), many children will have limited access to an assessment by a specialist. Furthermore, no screening is practiced at the out-patient departments or other clinics at present. In this context, even the children with high risk for ASD might not get referred to a specialist in the early stages. The proposed process will help in improving the diagnosis where the children in need of an early referral are identified through the pre-screening process using the Tree Model. This will ensure that specialist opinion is received for the children having high risk for ASD.

Furthermore, the classification tree can easily be automated in various platforms such as mobile or web platforms to automate the screening process. An interface can be designed to link the screening result with the diagnosis from the doctors/specialist from which the model can be re-evaluated. With more data and the feedback from doctors, the model can be further enhanced to increase the specificity. As this study was conducted in the clinical environment, we can use this automated screening tool in all the clinics throughout the country. This will ensure most of the children in the country are assessed for high risk in ASD in early years of their lives.

One of the problems we encountered in conducting this analysis was missing data for some predictors. This could be easily overcome if a digital form is introduced where default values would automatically be filled. In cases where this is not feasible, it is recommended to enforce entering an appropriate value for every variable as a mandatory requirement. The primary data source of this study was the clinical records of children, where the children were assessed by different psychiatrists and trained medical officers. As a result, the opinion of the psychiatrists and trained medical officers might have affected the final assessment. However, the effect could not be accounted for as the identification of the doctor, or the trained officer was not recorded. Therefore, we propose LRH to record the child's assessor in future records, so that their effect on the assessment could be considered through a mixed-effects logistic regression approach.

Due to limited data availability, in this study we could not do detailed analyses on age dependent factors. For example, in the form used at LRH, under developmental measures, there are "Number of functional words" and "Two-word sentences." However, as these two observations are age dependent, they were not considered for this analysis. Moreover, according to Table 1, 75% of children in this study are under 38 months. Though this is better in terms of early diagnosis, the study should be extended to cover a wider age range. Once a more complete data set is available, we would extend this study to complete age wise analysis.

ACKNOWLEDGMENTS

Grant Sponsor: Accelerating Higher Education Expansion and Development (AHEAD) Operation of the Ministry of Higher Education of Sri Lanka funded by the World Bank (https://ahead.lk/result-area-3/).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Hashan Peiris https://orcid.org/0000-0002-4721-9881
Darshana Chitraka Wickramarachchi https://orcid.org/0000-0001-6958-9667
Pradeepa Samarasinghe https://orcid.org/0000-0001-6908-3017
Philip Vance https://orcid.org/0000-0002-2189-3608
Dulangi Dahanayake https://orcid.org/0000-0002-1740-2183

Veerandi Kulasekara https://orcid.org/0000-0001-7079-3841
Madhuka Nadeeshani https://orcid.org/0000-0002-7312-1984

REFERENCES

Allison, P. D. (2001). *Missing data*. Sage Publications.

American Psychiatric Association. (2013). *DSM-5*. American Psychiatric Association. https://www.psychiatry.org/psychiatrists/practice/dsm

Arora, N. K., Nair, M., Gulati, S., Deshmukh, V., Mohapatra, A., Mishra, D., Patel, V., Pandey, R. M., Das, B. C., Divan, G., Murthy, G. V. S., Sharma, T. D., Sapra, S., Aneja, S., Juneja, M., Reddy, S. K., Suman, P., Mukherjee, S. B., Dasgupta, R., … Vajaratkar, V. (2018). Neurodevelopmental disorders in children aged 2–9 years: Population-based burden estimates across five regions in India. *PLoS Medicine*, *15*(7), e1002615.

Bhavnani, S., Lockwood Estrin, G., Arora, R., Kumar, D., Kakra, M., Vajaratkar, V., Juneja, M., Gulati, S., Patel, V., Green, J., & Divan, G. (2021). "I was confused… and still am" barriers impacting the help-seeking pathway for an autism diagnosis in urban North India: A mixed methods study. *Journal of Autism and Developmental Disorders*, *52*(4), 1778–1788.

Birx, H., Campbell, C., Rowe, C., Gjedde, A., Bauer, W. R., Wong, D. F., & American Psychiatric Association (2011). *DSM V TR* (Vol. 52, No. 1).

Bramer, M. (2007). *Principles of data mining* (Vol. 180). Springer.

Carruthers, S., Kinnaird, E., Rudra, A., Smith, P., Allison, C., Auyeung, B., Chakrabarti, B., Wakabayashi, A., Baron-Cohen, S., Bakolis, I., & Hoekstra, R. A. (2018). A cross-cultural study of autistic traits across India, Japan and the UK. *Molecular Autism*, *9*(1), 52. https://doi.org/10.1186/s13229-018-0235-3

Chandradasa, M., & Kuruppuarachchi, K. (2017). Child and youth mental health in post-war Sri Lanka. *BJPsych International*, *14*(2), 36–37.

Chlebowski, C., Green, J. A., Barton, M. L., & Fein, D. (2010). Using the childhood autism rating scale to diagnose autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*(7), 787–799.

Dahanayake, D., Rohanachandra, Y., & Wijetunge, G. (2015). *Factors affecting age at presentation of autism spectrum disorders: A descriptive survey from a child mental health clinic at lady ridgeway hospital*.

de Leeuw, A., Happ'E, F., & Hoekstra, R. A. (2020). A conceptual framework for understanding the cultural and contextual factors on autism across the globe. *Autism Research*, *13*(7), 1029–1050.

Fombonne, E. (2018). Editorial: The rising prevalence of autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *59*(7), 717–720. https://doi.org/10.1111/jcpp.12941

Frank, E. (2000). *Pruning decision trees and lists* (Unpublished doctoral dissertation). Citeseer.

Lady Ridgeway Hospital [LRH]. (2020). *Who we are – Lady Ridgeway Hospital for children*. LRH. http://lrh.health.gov.lk/who-we-are/

Maenner, M. J., Shaw, K. A., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., Daniels, J., Durkin, M. S., Fitzgerald, R. T., Kurzius-Spencer, M., Lee, L. C., Pettygrove, S., Robinson, C., Schulz, E., Wells, C., Wingate, M. S., Zahorodny, W., Yeargin-Allsopp, M., & Centers for Disease Control and Prevention (CDC). (2020). Prevalence of autism Spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries*, *69*(4), 1–12.

Martens, H. A., & Dardenne, P. (1998). Validation and verification of regression in small data sets. *Chemometrics and Intelligent Laboratory Systems*, *44*(1–2), 99–121.

Ministry of Health Nutrition and Indigenous Medicine. (2016). *Annual health statistics Sri Lanka 2016*. Ministry of Health Nutrition and Indigenous Medicine. http://www.health.gov.lk/moh_final/english/others.php?pid=110

Pace, N. L., & Briggs, W. M. (2009). Stepwise logistic regression. *Anesthesia & Analgesia*, *109*(1), 285–286.

Perera, H., Jeewandara, K. C., Guruge, C., & Seneviratne, S. (2013). Presenting symptoms of autism in Sri Lanka: Analysis of a clinical cohort. *Sri Lanka Journal of Child Health*, *42*(3), 139–143. https://doi.org/10.4038/sljch.v42i3.6017

Perera, H., Jeewandara, K. C., Seneviratne, S., & Guruge, C. (2016). Outcome of home-based early intervention for autism in Sri Lanka: Follow-up of a cohort and comparison with a nonintervention group. *BioMed Research International*, *2016*, 3284087. https://doi.org/10.1155/2016/3284087

Perera, H., Jeewandara, K. C., Seneviratne, S., & Guruge, C. (2017). Culturally adapted pictorial screening tool for autism spectrum disorder: A new approach. *World Journal of Clinical Pediatrics*, *6*(1), 45. https://doi.org/10.5409/wjcp.v6.i1.45

Perera, H., Wijewardena, K., & Aluthwelage, R. (2009). Screening of 18-24-month-old children for autism in a semi-urban community in Sri Lanka. *Journal of Tropical Pediatrics*, *55*(6), 402–405. https://doi.org/10.1093/tropej/fmp031

Petrocchi, S., Levante, A., & Lecciso, F. (2020). Systematic review of level 1 and level 2 screening tools for autism spectrum disorders in toddlers. *Brain Sciences*, *10*(3), 180. https://doi.org/10.3390/brainsci10030180

Postgraduate Institute of Medicine (PGIM) Board Certified Specialists – Child & Adolescent Psychiatry (n.d.). http://online.pgim.cmb.ac.lk/bocfinal/boc/

Roberts, J. M. A. (1989). Echolalia and comprehension in autistic children. *Journal of Autism and Developmental Disorders*, *19*(2), 271–281.

Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, *3*, 153–160.

Shouman, M., Turner, T., & Stocker, R. (2011). Using decision tree for diagnosing heart disease patients. In *Proceedings of the ninth Australasian data mining conference* (Vol. 121, pp. 23–30).

Thabtah, F., & Peebles, D. (2019). Early autism screening: A comprehensive review. *International Journal of Environmental Research and Public Health*, *16*(18), 3502. https://www.mdpi.com/1660-4601/16/18/3502

Western Psychological Services. (2012). *(ADOS®-2) autism diagnostic observation schedule, second edition—WPS*. Western Psychological Services. https://www.wpspublish.com/ados-2-autism-diagnostic-observation-schedule-second-edition

World Health Organization [WHO]. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. World Health Organization.

World Health Organization [WHO]. (2020). *ICD-11 international classification of diseases for mortality and morbidity statistics. 2018*. World Health Organization.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Peiris, H., Wickramarachchi, D. C., Samarasinghe, P., Vance, P., Dahanayake, D., Kulasekara, V., & Nadeeshani, M. (2022). Diagnosing autism in low-income countries: Clinical record-based analysis in Sri Lanka. *Autism Research*, 1–10. https://doi.org/10.1002/aur.2765

View publication stats