Accepted manuscript for:

Author/s: Md Delowar Hossain; Ashad Kabir; Adnan Anwar and Zahid Islam

Title: Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets

Journal: Health Information Science and Systems

ISSN: 2047-2501

Year: 2021

Volume: 9

Pages: 1-13

Article Number: 17

Abstract: Autism Spectrum Disorder (ASD), which is a neuro development disorder, is often accompanied by sensory issues such an over sensitivity or under sensitivity to sounds and smells or touch. Although its main cause is genetics in nature, early detection and treatment can help to improve the conditions. In recent years, machine learning based intelligent diagnosis has been evolved to complement the traditional clinical methods which can be time consuming and expensive. The focus of this paper is to find out the most significant traits and automate the diagnosis process using available classification techniques for improved diagnosis purpose. We have analyzed ASD datasets of toddler, child, adolescent and adult. We have evaluated state-of-the-art classification and feature selection techniques to determine the best performing classifier and feature set, respectively, for these four ASD datasets. Our experimental results show that multilayer perceptron (MLP) classifier outperforms among all other benchmark classification techniques and achieves 100% accuracy with minimal number of attributes for toddler, child, adolescent and adult datasets. We also identify that 'relief F' feature selection technique works best for all four ASD datasets to rank the most significant attributes.

DOI to published version: http://dx.doi.org/10.1007/s13755-021-00145-9

Noname manuscript No.
(will be inserted by the editor)

Detecting Autism Spectrum Disorder using Machine Learning Techniques

An Experimental Analysis on Toddler, Child, Adolescent and Adult Datasets

Md Delowar Hossain · Muhammad Ashad Kabir · Adnan Anwar · Md Zahidul Islam

Received: date / Accepted: date

Abstract Autism Spectrum Disorder (ASD), which is a neuro development disorder, is often accompanied by sensory issues such an over sensitivity or under sensitivity to sounds and smells or touch. Although its main cause is genetics in nature, early detection and treatment can help to improve the conditions. In recent years, machine learning based intelligent diagnosis has been evolved to complement the traditional clinical methods which can be time consuming and expensive. The focus of this paper is to find out the most significant traits and automate the diagnosis process using available classification techniques for improved diagnosis purpose. We have analyzed ASD datasets of toddler, child, adolescent and adult. We have evaluated state-of-the-art classification and feature selection techniques to determine the best performing classifier and feature set, respectively, for these four ASD datasets. Our experimental results show that multilayer perceptron (MLP) classifier outperforms among all other benchmark classification techniques and achieves 100% accuracy with minimal number of attributes for toddler, child, adolescent and adult datasets. We also identify that 'relief F' feature selection technique works best for all four ASD datasets to rank the most significant attributes.

Keywords Autism spectrum disorder · machine learning · feature selection · classification · ASD detection

M D Hossain
School of Computing and Mathematics
Charles Sturt University
NSW, Australia
E-mail: mailtodelowar@gmail.com

M A Kabir (corresponding author)
School of Computing and Mathematics
Charles Sturt University
NSW, Australia
Tel: +61263386259
E-mail: akabir@csu.edu.au

A Anwar
School of Information Technology
Deakin University
Waurn Ponds, Australia
E-mail: adnan.anwar@deakin.edu.au

M Z Islam
School of Computing and Mathematics
Charles Sturt University
NSW, Australia
E-mail: zislam@csu.edu.au

1 Introduction

Autism spectrum disorder (ASD), is a neurological developmental disorder. It affects how people communicate and interact with others, as well as how they behave and learn [1]. The symptoms and signs appear when a child is very young. It is a lifelong condition and cannot be completely cured. A study found that 33% of children with difficulties other than ASD have some ASD symptoms while not meeting the full classification criteria [2].

ASD has a significant economic impact both due to the increase in the number of ASD cases worldwide, and the time and costs involved in diagnosing a patient. Early detection of ASD can help both patient and healthcare service providers by prescribing proper therapy and/or medication needed and thereby reducing the long-term costs associated with delayed diagnosis. On the other hand the traditional clinical methods such as Autism Diagnostic Interview Revised (ADI-R) and Autism Diagnostic Observation Schedule Revised (ADOS-R), are time consuming and cumbersome [3,4]. The child who are too young and has delayed speech issue roughly score 25% of the total ADI-R items because the verbal sections cannot be answered accurately for the patient. Besides, conducting interview with a caregiver by a trained examiner takes 90 to 150 minutes which is cumbersome and often misses data. On the other hand, the detection of ASD by ADOS-R depends on measurements of the scoring based on the answers provided. Moreover, one of the major disadvantages of this approach is the tendency to over classify children who have other clinical disorders [5]. So, the healthcare professionals are in urgent need of a time efficient, easy and accessible ASD screening method that can accurately detect whether a patient with a certain measured characteristic has ASD and can inform individuals whether they should pursue a formal clinical diagnosis. Presently, the available datasets are few and associated with clinical diagnosis which is mostly genetic in nature, e.g., AGRE [6], National Database of Autism Research (NDAR) [7] and Boston Autism Consortium (AC) [8].

Now a days, machine learning has been applied to detect various diseases including depression [9] and ASD [7, 10, 11]. The primary objectives of applying machine learning techniques are to improve diagnosis accuracy and reduce diagnosis time of a case in order to provide quicker access to health care services. Since the diagnosis process of a case involves coming up with the right class (ASD, No-ASD) based on the input case features, this process can be attributed as a classification task in machine learning. In this paper, we apply various classification techniques to obtain improved accuracy on the results of detecting ASD cases for all four datasets. The main contributions of this paper can be summarised as follows:

1. We analyze the attributes of Toddler, Child, Adolescent and Adult ASD datasets, and identify associations between the demographic information and ASD cases.
2. We explore benchmark feature selection methods and identify the method that performs best for all four ASD datasets to select the set of most significant features to achieve highest classification accuracy. Our analysis shows that appropriate feature selection can significantly improve the ASD classification performance.
3. We compare state-of-the-art classification techniques and identify the best performing classifier for all four ASD datasets.

The rest of the paper is organized as follows. Section 2 reviews related work. In Section 3, we present our methodology. Description of the datasets, preprocessing and exploratory analysis are presented in Section 4. The performance comparison of benchmark classification techniques is presented in Section 5. Section 6 reports feature engineering results. After a thorough comparison of our approach with existing approaches in Section 7, we conclude the paper in Section 8.

2 Related work

A number of research have adopted machine learning techniques to improve the diagnosis process of ASD [7, 10, 11]. The primary motivation of using machine learning models for ASD is to reduce the early detection time that enables quicker access to health care services and improves diagnostic accuracy [12]. We can categorize the ASD detection study in two groups – Video clip-based study and AQ-based study.

2.1 Video clip-based study

Tariq et al. [13] have hypothesized that the use of machine learning analysis on home video can speed up the diagnosis time without compromising accuracy. Authors have analyzed item-level records from two standard diagnostic instruments to construct machine learning classifiers optimized for sparsity, interpretability and accuracy. Authors have considered eight machine learning models to apply on 162 two-minutes home videos of children with and without ASD. Besides, a mobile web portal has been created for video raters to assess 30 behavioral features (e.g., eye contact, social smile, etc.) that are used by eight different machine learning models for detecting ASD. The result shows that 94% accuracy is achieved for each case using cross-validation testing and subsequent independent validation from previous work. However, this method is also time consuming since the video needs to be recorded and assessed for the rating based on 30 questions. Whereas we adopt a method that only uses mobile app [14] from where users can easily select the appropriate answers of the ten ASD questions. Besides, improved analysis based on reduced number of attributes can significantly improve the performance of the ASD detection.

Andrea et al. [15] analyzed video gesture for detecting ASD. Authors have devised an experimental setup by recording video of patient and healthy children performing simple gesture of grasping a bottle. By only processing the video clips depicting the grasping action using a recurrent deep neural network, they are able to classify ASD and non-ASD cases with a good accuracy. In that work, authors followed a common procedure where each video is split into 15-frame clips and passed through the entire model that outputs a binary vector containing, for each frame, the probability of ASD and No-ASD. Each clip is considered independent during the training. The test accuracy for each subject is computed by averaging the scored probabilities over all the frames of a video. But the model decreases its effectiveness after a threshold of 0.9. The results support the hypothesis that feature tagging of home videos using machine learning classification of autism can yield accurate outcomes in short time frames.

2.2 AQ-based study

Autistic-spectrum Quotient (AQ) is a screening method developed by Baron-Cohen et al. [16]. Later Allison et al. [10] proposes the AQ 10-adult and AQ 10-child, shortened versions of the original AQ. This questionnaire based attempt is said to increase the efficiency of ASD screening.

There are two versions of the available ASD datasets – version-1 (v1) [17] that has 20 attributes and version-2 (v2) [18] that has 23 attributes and more records than version-1. Version-1 dataset for Toddler was not available. Almost all of the previous research works were based on the version-1 ASD dataset. The Autism questionnaires (AQ1 to AQ10) remain same for the both versions.

Kanad Basu [19] has analyzed the version-1 Adult ASD dataset using supervised machine learning techniques such as decision tree, random forest, support vector machines (SVM), k-nearest neighbor (kNN), naïve bayes, logistic regression, linear discriminant analysis (LDA) and multilayer perceptron (MLP). Author concludes that SVM classification technique scores the highest accuracy among all other benchmark techniques for v1 adult ASD dataset. Later McNamara et al. [20] also classify the same dataset by applying Decision Tree and random forest classifiers considering improved data pre-processing where authors remove least significant attributes and records with missing values before applying the classifiers. The comparison results between these two classifiers show that the random forest results more accuracy for version-1 adult ASD dataset. Hossain et al. [21] conducted the research on v1 child ASD dataset following the same methodology and applied 27 benchmark classifiers. They have also found that the sequential minimal optimization (SMO) classifier performs best on detecting child ASD cases. Beside this, they also identified the dominant features in detecting child autism. Raj et al. [22], Baranwal et al. [23] and Erkan et.al. [24] used v1 adult, adolescent and child ASD dataset and applied machine learning techniques to detect ASD. They have used all 20 attributes in classification.

To the best of our knowledge, only Thabtah et al. [25] used v2 three ASD datasets (Child, Adolescent and Adult) and have applied a rule-based machine learning technique for classification. In contrast, we have used all four v2 ASD datasets (Toddler, Child, Adolescent and Adult) and applied 27 benchmark classification techniques to identify the best classifier. To enhance classification accuracy, we further apply feature engineering approach to identify minimal set of significant features/attributes and use them in classification (instead of using all attributes).

3 Methodology

Figure 1 depicts an overview of our methodology for detecting ASD cases, briefly described below.

- **Step 1: Data preprocessing and analysis.** In this step, we perform a detailed data-preprocessing before going into detailed analysis and classification. The ASD datasets [18] contain few records with missing values. Also there are some attributes which represent meta-information (e.g., used app before, who completed the test) and not related to ASD. Thus, the datasets need to be cleaned/preprocessed before applying classification. Section 4 presents details of our data preprocessing and analysis including dataset description, data cleansing/preprocessing, and statistical analysis of association between attributes.

- **Step 2: Apply benchmark classification techniques.** It this step, we apply various classification techniques. In this study, we have applied 27 classification techniques to all four ASD datasets and evaluated their performance using accuracy and F-measure through 10-fold cross validation. Finally, we select top eight classifiers for further evaluation. Section 5 reports the results.

- **Step 3: Feature engineering.** Classification accuracy may degrade if all attributes in a dataset are used (for instance, see Table 8 results for adolescent dataset). Furthermore, less number of attributes reduces the required resources (time and memory) for training a model. Thus, in this step, we rank attributes/features to identify the optimal set of most significant attributes that results highest accuracy. We apply five benchmark feature ranking techniques to compare and identify the technique that ranked attributes consistently across all four ASD datasets

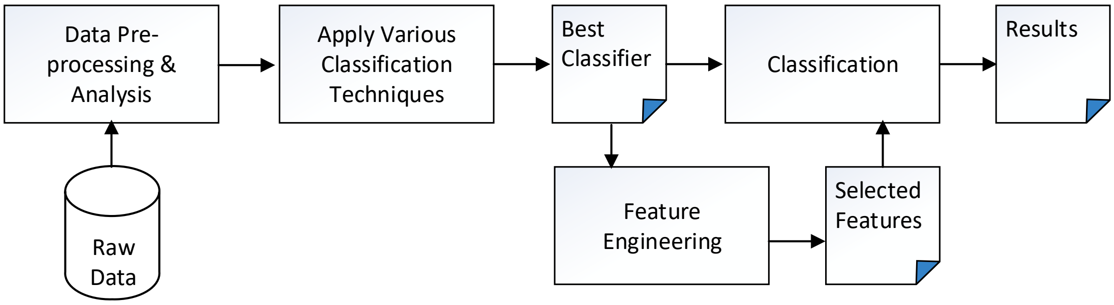


Fig. 1: Methodology for detecting autism

and can be used to identify the optimal set of most significant attributes. Finally, we identify the best classifier among the top eight classifiers (selected in the previous step) that scores highest accuracy for the optimal set of attributes. Section 6 describes this step in details with our findings.

– **Step 4: Classification results comparison.** ASD positive result from the classification indicates that the patient needs to undergo further medical diagnosis and necessary treatments. Therefore, the classification accuracy is important to reduce the false positive and overall improvement of the outcome. Section 7 presents a comparative analysis of our work with state-of-the-art results. The results show that multilayer perceptron (MLP) classifier with 'relief F' feature engineering achieves 100% accuracy using top ten attributes for all four ASD datasets, and outperforms state-of-the-art results.

4 Data preprocessing and analysis

4.1 Dataset description

Our used ASD datasets (version 2) [18] mainly comprise 23 attributes (except the toddler dataset which has 18 attributes). The attribute descriptions are given in Table 1. All datasets have ten binary attributes representing the answers of screening questions (A1 to A10) as well as the categorical variables such as gender, ethnicity, jaundice, family_ASD, residence and ASD class. Thee datasets also have two numeric variables such as age and screen score/results. We found that five attributes were absent in toddler ASD dataset: who completed the test (user), why taken the screening, used_app_before, country of residence and language spoken.

Child and adolescent datasets have same screening questions (A1 to A10) whereas toddler and adult have some unique questions. We have furnished all the questionnaires of ASD dataset according to the sequence for child, adolescent, adult and toddler (please see the Description column for details) in Table 1. The class value is assigned during the process of data collection based on the answers of AQ-10 (A1 to A10) questions. The class value "No" is assigned when the final score of AQ-10 methods scores less than or equal to 7. Otherwise, it is assigned "Yes" which indicates that the individual does have ASD. However, for the toddler dataset the cut-off score is less than or equal to 4. So, in this case if the total score ≥ 4, it is considered that the subject has ASD.

Table 2 reports number of ASD and non-ASD cases in the datasets. Figure 2a shows gender-wise class distribution of the considered four ASD datasets. Here, we observe that the child and adolescent datasets are balanced but toddler and adult datasets are not balanced considering the total number of ASD cases and/or gender distribution.

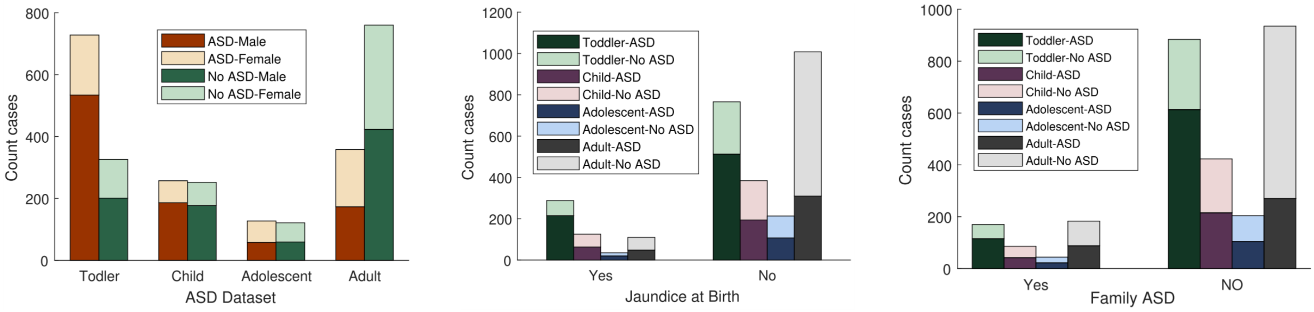
4.2 Data preprocessing

In order to simplify our model and to improve classification accuracy, we clean up datasets by removing the instances with missing values. Afterwards, we preprocess the datasets by reducing the attributes those are meta information and are not associated with ASD, listed below:

- Case
- Used App Before
- User (who completed the screening)
- Language
- Why taken the screening
- Age Description

Table 1: ASD Dataset Description

Attribute	**Type**	**Description**
Age	Number	Age in months/years
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (yes or no)	Whether the case was born with jaundice
Family member with PDD	Boolean (yes or no)	Whether any immediate family member has a PDD
Who is completing the test (User)	String	Parent, self, caregiver, medical staff, clinician,etc.
Why taken the screening	Meta	The person can write short reason for completing the task
Used_App_Before	Boolean (yes or no)	This answer would be binary
Language spoken	String	The user will give his/her native language information
Country of residence	String	List of countries in text format
Used the screening app before	Boolean (yes or no)	Whether the user has used a screening app
Screening Method Type	Integer(0,1,2,3)	The type of Screening Methods choses based on age category (0=toddler, 1=Child, 2= Adolescent, 3= Adult)
Question 1 (A1)	Binary (0, 1)	S/he often notices small sounds when others do not, (Child, Adolescent) S/he notices patterns in things all the time, (Adult) Does your child look at you when you call his/her name? (Toddler)
Question 2 (A2)	Binary (0, 1)	S/he usually concentrates more on the whole picture, rather than the small detail, (child, Adolescent, Adults) How easy is it for you to get eye contact with your child? (Toddler)
Question 3 (A3)	Binary (0, 1)	In a social group, s/he can easily keep track of several different people's conversations, (child, Adolescent) I find it easy to do more than one thing at once, (Adult) Does your child point to indicate that s/he wants something? (e.g. a toy that is out of reach) (Toddler)
Question 4 (A4)	Binary (0, 1)	S/he finds it easy to go back and forth between different activities, (child, Adolescent) If there is an interruption, s/he can switch back to what s/he was doing very quick, (Adult) Does your child point to share interest with you? (e.g. pointing at an interesting sight) (Toddler)
Question 5 (A5)	Binary (0, 1)	S/he doesn't know how to keep a conversation going with his/her peers, (child, Adolescent) I find it easy to read between the lines when someone is talking to me, (Adult) Does your child pretend? (e.g. care for dolls, talk on a toy phone) (Toddler)
Question 6 (A6)	Binary (0, 1)	S/he is good at social chit-chat, (child, Adolescent) I know how to tell if someone listening to me is getting bored, (Adult) Does your child follow where you're looking? (Toddler)
Question 7 (A7)	Binary (0, 1)	When s/he is read a story, s/he finds it difficult to work out the character's intentions or feelings, (Child) When s/he was younger, s/he used to enjoy playing games involving pretending with other children, (Adolescent) When I'm reading a story, I find it difficult to work out the characters' intentions, (Adult) If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging them (Toddler)
Question 8 (A8)	Binary (0, 1)	When s/he was in preschool, s/he used to enjoy playing games involving pretending with other children, (Child) S/he finds it difficult to imagine what it would be like to be someone else, (Adolescent) I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant etc), (Adult) Would you describe your child's first words as: (Toddler)
Question 9 (A9)	Binary (0, 1)	S/he finds it easy to work out what someone is thinking or feeling just by looking at their face, (Child) S/he finds social situations easy, (Adolescent) I find it easy to work out what someone is thinking or feeling just by looking at their face, (Adult) Does your child use simple gestures? (e.g. wave goodbye) (Toddler)
Question 10 (A10)	Binary (0, 1)	S/he finds it hard to make new friends, (Child, Adolescent) I find it difficult to work out people's intentions, (Adult) Does your child stare at nothing with no apparent purpose? (Toddler)
Screening Score	Integer (0 to 10)	The final score obtained based on the scoring algorithm of the screening method used. This is computed in an automated manner
Class	Binary (0, 1)	Subject was diagnosed with ASD or not: 1 - ASD, 0 - Not ASD



(a) Gender-wise distribution of ASD cases

(b) Impact of 'jaundice at birth' on ASD

(c) Impact of 'family ASD' on their offspring ASD

Fig. 2: Analysis of toddler, child, adolescent and adult ASD datasets

Table 2: Number of ASD and non-ASD cases in the datasets (version 2)

Dataset	ASD	Non-ASD
Toddler	728	326
Child	257	252
Adolescent	127	122
Adult	358	760

- Screening Type
- Score

We observed that the 'score' attribute value 7 or higher classified as ASD_Class = YES for child, adolescent and adult ASD datasets, and the score value 4 or higher classified as ASD_Class = YES for toddler dataset. So, including this attribute in classification implies that the classification algorithm already has the outcome of the target variable. For this reason this attribute is removed during analysis. Finally, we select 16 attributes for child, adolescent and adult datasets and 15 attributes ("residence" attribute was absent) for toddler dataset.

In the next subsections, we investigate the impact/association of jaundice, family ASD and ethnicity on ASD cases for all four ASD datasets.

4.3 Association between jaundice at birth and ASD cases

Figure 2b shows the distribution of ASD cases by jaundice at birth for all four datasets. Our association analysis results in Table 3 show that there is no significant association between patients' jaundice at birth and ASD for child and adolescent in particular. Whereas the result for toddler and adult dataset have shown a significant association (as $p < .05$) but the association strength found very weak/negligible (as $V < .1$).

4.4 Association between family ASD and ASD cases

Figure 2c shows the distribution of ASD by Family ASD cases for all four datasets. This figure and the association analysis results in Table 3 show that there is no significant association between family ASD cases and their children ASD cases for toddler, child and adolescent. Whereas we got a significant p-value ($p < .05$) for adult dataset but the resulted Cramer's V value ($V < .3$) makes the strength of the association low.

4.5 Association between ethnicity and ASD cases

Figure 3 depicts the distribution of ASD cases by various ethnicity for all four datasets. This figure and the association analysis results in Table 3 show that there is a significant association between ethnicity and ASD cases for all four datasets ($p < .05$). The strength of the association found moderate for adolescent and adult whereas it was low for toddler and child.

5 Applying classification techniques

We apply 27 benchmark machine learning classification techniques to determine the best technique(s) that results highest accuracy. No single technique is universally perfect for different datasets and all types of classification problems. After introducing the used evaluation matrix in Section 5.1, we present our comparison results in Section 5.2.

5.1 Evaluation matrix

For a given dataset and a predictive model, every data point will lie on one of the below four categories.

Table 3: Measuring association using chi-square test

Categorical variables		Test value	p-value	Cramer's V	Interpretation of association
Jaundice at birth	Toddler	5.781*	.016	.074	negligible ($0 < V \leq .1$)
	Child	.001	.981	.001	no significant ($p \geq \alpha$)
	Adolescent	.574	.449	.048	no significant ($p \geq \alpha$)
	Adult	7.561*	.006	.082	negligible ($0 < V \leq .1$)
Family ASD	Toddler	.192	.661	.014	no significant ($p \geq \alpha$)
	Child	.113	.736	.015	no significant ($p \geq \alpha$)
	Adolescent	.031	.860	.011	no significant ($p \geq \alpha$)
	Adult	25.947*	.000	.152	low association ($.1 < V \leq .3$)
Ethnicity	Toddler	43.571*	.000	.203	low association ($.1 < V \leq .3$)
	Child	17.841*	.022	.187	low association ($.1 < V \leq .3$)
	Adolescent	29.174*	.000	.343	moderate association ($.3 < V \leq .5$)
	Adult	121.831*	.000	.330	moderate association ($.3 < V \leq .5$)

* indicates significant at 5% level, i.e., $\alpha = .05$

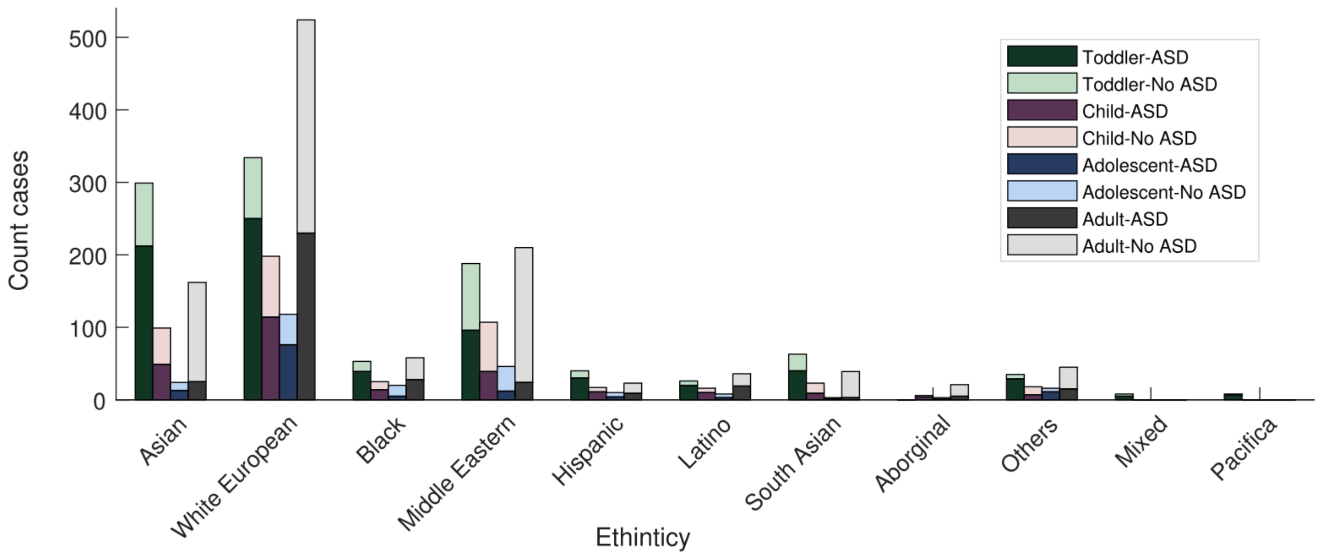


Fig. 3: Ethnicity in ASD datasets

- True Positive (TP): The individual having ASD and is correctly predicted as having ASD.
- True Negative (TN): The individual not having ASD and was correctly predicted as not having ASD.
- False Positive (FP): The individual not having ASD, is incorrectly predicted as having ASD.
- False Negative (FN): The individual having ASD, is incorrectly predicted as not having ASD.

Those categories are used to compute the following evaluation matrix:

Accuracy: It is the measure of correct predictions made by the classifier. Accuracy is the number of correctly identified predictions by total number of predictions:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Precision: It measures the accuracy of positive predictions. It is the ratio of true positive out of the total observed positive.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall/Sensitivity: This is also called true positive rate. It is the proportion of samples that are genuinely positive by all positive results obtained during the test.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F-Measure: The F-score (or F-measure) considers both the precision and the recall of the test to compute the score. The traditional or balanced F-score (F1 score) is the harmonic mean of the precision and recall:

$$F - Measure\ (F1) = \frac{2 \times Precision}{Precision + Recall} \tag{4}$$

Table 4: Comparison of classification techniques

Classifier	Toddler		Child		Adolescent		Adult	
	F1	Acc. (%)	F1	Acc. (%)	F1	Acc. (%)	F1	Acc. (%)
OneR	0.81	80.551	0.79	78.98	0.76	76.63	0.83	83.18
PART	0.95	94.59	0.90	90.37	0.87	87.10	0.94	94.45
JRip (RIPPER)	0.94	93.45	0.87	87.43	0.85	84.68	0.94	94.28
Ridor	0.92	92.40	0.88	87.82	0.85	85.08	0.93	92.94
Nneg	0.93	92.60	0.82	82.32	0.81	81.05	0.88	88.55
NaiveBayes	0.96	95.54	0.93	92.93	0.91	91.13	0.94	94.10
LibSVM	0.97	96.96	0.49	49.31	0.93	92.74	0.96	96.42
Multilayer Perceptron (MLP)	1.0	100	1.0	100	0.98	98.79	1.0	100
Logistic regression (LR)	0.97	99.62	0.99	99.21	0.95	94.76	0.97	97.32
Simple logistic (SL)	1.0	100	0.99	99.80	0.96	95.97	0.99	99.82
SMO	1.0	100	1.0	100	0.97	97.18	1.0	100
IBK	0.93	92.70	0.88	88.016	0.89	88.71	0.92	92.13
KStar	0.95	94.50	0.84	84.09	0.88	88.31	0.93	92.58
LWL	0.85	84.54	0.79	78.98	0.79	79.03	0.79	78.35
Bagging	0.93	93.17	0.80	80.16	0.76	76.21	0.87	87.30
Iterative classifier optimizer (ICO)	1.0	100	0.99	99.41	0.97	97.18	0.99	99.91
LogitBoost (LB)	1.0	100	0.99	99.41	0.97	97.18	0.99	99.91
Multi class classifier	0.99	99.62	0.99	99.21	0.95	94.76	0.97	97.32
Real Adaboost (RAB)	1.0	100	0.99	99.80	0.97	97.18	0.99	99.82
Hoeffding Tree	0.95	95.25	0.91	91.16	0.90	90.73	0.93	92.84
J48 (C4.5)	0.90	90.32	0.89	88.99	0.81	81.85	0.93	93.29
LMT	1.0	100	0.99	99.80	0.96	95.96	0.99	99.82
NBTree	0.96	95.54	0.93	93.12	0.87	86.69	0.94	94.19
Random Forest	0.95	95.45	0.88	87.63	0.88	87.5	0.92	92.13
Random Tree	0.91	91.08	0.79	78.59	0.79	79.84	0.86	85.87
Simple CART	0.91	90.89	0.81	81.34	0.81	81.05	0.90	89.62
SysFor	0.93	92.79	0.87	86.84	0.86	85.89	0.93	93.38

Note: Acc. indicates Accuracy

5.2 Comparison of classification techniques

In classification, we consider all 15 attributes for toddler dataset and 16 attributes for child, adolescent and adult datasets. Table 4 presents classification results (F-Measure and accuracy) of 27 benchmark classification techniques for four ASD datasets. The results show that 8 out of 27 classifiers such as SMO, logistic regression, multi class classifier, simple logistic, logit boost, iterative classifier optimizer, real adaboost, LMT and multilayer perceptron (MLP) demonstrate competitive performance as they result 100% accuracy at least for one of the ASD datasets.

In the next section, we apply feature engineering techniques to further enhance the classification accuracy through selecting the most significant minimal set of attributes and identifying the classifier(s) that outperforms among the above eight classifiers for all four ASD datasets.

6 Feature Engineering

We have applied and compared five prominent feature selections methods such as Information gain, Chi-square test, Pearson correlation, One-R and Releif F on four ASD datasets and listed their attributes' ranking side by side (see Table 5).

6.1 Analyzing feature selection techniques and feature ranking

In this section, we explore the effectiveness of feature selection techniques on ASD datasets. We know that the attributes which are the answers of the question A1 to A10, mainly the deciding factors of the ASD cases. Besides, answers of the demographic questions have little to no effect for identifying ASD cases. So, by comparing the five feature selection methods, we find that Relief F attribute selection method performs best among the five methods and it is capable of ranking A1 to A10 attributes before the demographic attributes for all four ASD datasets.

We count the total number of occurrences of the attributes in four ASD datasets respectively and compare the effects of the attributes in detecting ASD cases. In Figure 4a, 4b, 4c and 4d, the first column in each group represents the score '1' of the attribute when the ASD case is "yes" (lower portion) and "no" is the upper por-

Table 5: Comparison of feature selection methods

Rank	Information Gain				Chi Squared				Correlation				One R				**Relief F**			
	Toddler	Child	Adolescent	Adult	Toddler	Child	Adolescent	Adult	Toddler	Child	Adolescent	Adult	Toddler	Child	Adolescent	Adult	Toddler	Child	Adolescent	Adult
1	A9	A4	A6	A6	A9	A4	A6	A6	A9	A4	A6	A6	A7	A4	A6	A6	A9	A4	A6	A5
2	A5	Res	Res	A5	A6	A6	A3	A9	A6	A6	A3	A9	A6	A9	A3	A9	A5	A8	A3	A6
3	A6	A6	A3	A9	A5	A9	Res	A5	A5	A9	A4	A5	A5	A8	A4	A5	A2	A9	A5	A9
4	A7	A9	A4	Res	A7	Res	A4	A4	A7	Res	A5	A4	A9	A6	A5	A4	A6	A10	A4	A3
5	A4	A8	A5	A4	A4	A8	A5	Res	A4	A8	A9	A3	A1	A5	A9	A3	A7	A1	A7	A4
6	A1	A5	A9	A3	A1	A5	A9	A3	A1	A5	A10	A10	A4	A10	A10	A7	A4	A5	A9	A10
7	A2	A3	A10	A10	A2	A3	A10	A10	A2	A3	A7	A7	Age	A3	A7	Res	A1	A6	A10	A7
8	A8	A10	A7	A7	A8	A10	A7	A7	A8	A10	A2	A2	A3	A1	A2	A2	A8	A3	A2	A1
9	A3	A1	Ethn	Ethn	A3	A1	Ethn	Ethn	A3	A1	A1	A1	FASD	A7	A1	A8	A3	A7	A1	A8
10	Ethn	A7	A2	A1	Ethn	A7	A2	A2	A10	A7	A8	A8	Jaun	A2	A8	A1	A10	A2	A8	A2
11	A10	A2	A1	A2	A10	A2	A1	A1	Sex	A2	Ethn	Ethn	A10	Res	Res	A10	Ethn	Res	Ethn	Res
12	Age	Ethn	A8	A8	Age	Ethn	A8	A8	Jand	Ethn	Res	FASD	A8	Ethn	Ethn	Jand	Jand	Age	Res	Ethn
13	Sex	Sex	Jand	FASD	Sex	Sex	Jand	FASD	Ethn	Sex	Age	Res	Sex	Age	Age	Sex	Age	Sex	Age	Jand
14	Jand	FASD	Sex	Jand	Jand	FASD	Sex	Jand	Age	FASD	Jand	Jand	A2	Sex	Jand	Ethn	FASD	Jand	FASD	Age
15	FASD	Jand	FASD	Sex	FASD	Jand	FASD	Sex	FASD	Jand	Sex	Age	Ethn	Jand	Sex	FASD	Sex	FASD	Sex	Sex
16	–	Age	Age	Age	–	Age	Age	Age	–	Age	FASD	Sex	–	FASD	FASD	Age	–	Ethn	Jand	FASD

NOTE: 'Res' means Residence, 'Ethn' means Ethnicity, 'FASD' means Family ASD, 'Jand' means Jaundice

Table 6: Classifiers' accuracy (%) with increasing number of attributes according to their ranks (from Relief F) for toddler ASD dataset

Attribute set	Size	MLP	SMO	LR	SL	LB	ICO	RAB	LMT
{A9}	1	76.28	76.28	76.28	76.28	76.28	76.28	76.29	76.29
{A9,A5}	2	85.96	85.96	85.96	85.96	85.96	85.96	85.96	85.96
{A9,A5,A2}	3	88.99	88.99	88.99	88.99	88.99	88.99	89.0	89.0
{A9,...,A6}	4	88.43	88.43	89.18	89.18	88.99	89.37	89.38	89.19
{A9,...,A7}	5	91.94	92.32	92.32	92.32	92.32	92.32	92.32	92.32
{A9,...,A4}	6	91.08	91.56	91.08	91.46	91.56	91.65	91.75	91.18
{A9,...,A1}	7	93.17	94.50	94.50	94.50	94.31	94.31	94.03	94.5
{A9,...,A8}	8	94.78	96.39	95.83	96.02	96.11	96.39	96.12	96.02
{A9,...,A3}	9	96.02	96.02	95.64	96.02	96.02	96.02	95.93	96.02
{A9,...,A10}	10	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
{A9,...,Ethn}	11	100	100	99.72	100	100	100	100	100
{A9,...,Jand}	12	100	100	99.62	100	100	100	100	100
{A9,...,Age}	13	100	100	99.62	100	100	100	100	100
{A9,...,FASD}	14	100	100	99.53	100	100	100	100	100
{A9,...,Sex}	15	100	100	99.62	100	100	100	100	100

Table 7: Classifiers' accuracy (%) with increasing number of attributes according to their ranks (from Relief F) for child ASD dataset

Attribute set	Size	MLP	SMO	LR	SL	LB	ICO	RAB	LMT
{A4}	1	78.98	78.98	78.98	78.98	78.98	78.98	78.98	78.98
{A4,A8}	2	78.98	78.98	78.98	78.98	78.98	78.98	78.98	78.98
{A4,A8,A9}	3	80.55	78.78	81.14	81.14	81.14	81.14	80.56	81.14
{A4,...,A10}	4	81.73	82.51	83.50	83.50	83.50	83.50	83.5	82.72
{A4,...,A1}	5	83.69	87.43	85.66	86.44	86.05	85.46	85.66	85.47
{A4,...,A5}	6	91.75	90.37	88.80	89.78	89.19	89.00	89.79	92.15
{A4,...,A6}	7	92.73	90.96	91.16	91.16	90.77	90.57	91.31	91.56
{A4,...,A3}	8	93.52	91.75	93.52	93.52	93.52	92.93	93.13	93.52
{A4,...,A7}	9	95.68	93.71	94.50	93.91	94.50	94.30	94.9	93.91
{A4,...,A2}	10	**100**	**100**	**100**	99.80	99.41	99.41	99.81	99.81
{A4,...,Res}	11	100	100	97.25	99.80	99.41	99.41	99.81	99.81
{A4,...,Age}	12	100	100	97.25	99.80	99.41	99.41	99.81	99.81
{A4,...,Sex}	13	100	100	97.25	99.80	99.41	99.41	99.81	99.81
{A4,...,Jand}	14	100	100	97.45	99.80	99.41	99.41	99.81	99.81
{A4,...,FASD}	15	100	100	97.25	99.80	99.41	99.41	99.81	99.81
{A4,...,Ethn}	16	100	100	98.82	99.80	99.41	99.41	99.81	99.81

tion of the column, and the second column represents the score '0' of the attribute when the ASD case is "yes" represents the lower portion and "no" represents the upper portion. We compare if the attribute has a better ratio in detecting ASD "yes" and "no" cases in the first and second column respectively. Thus, if we rank these attributes, we find that it is consistent with Relief F feature ranking for all four datasets. Thus, we use Relief F feature selection method to identify the minimal set of most significant features that provides best accuracy.

Table 8: Classifiers' accuracy (%) with increasing number of attributes according to their ranks (from Relief F) for adolescent ASD dataset

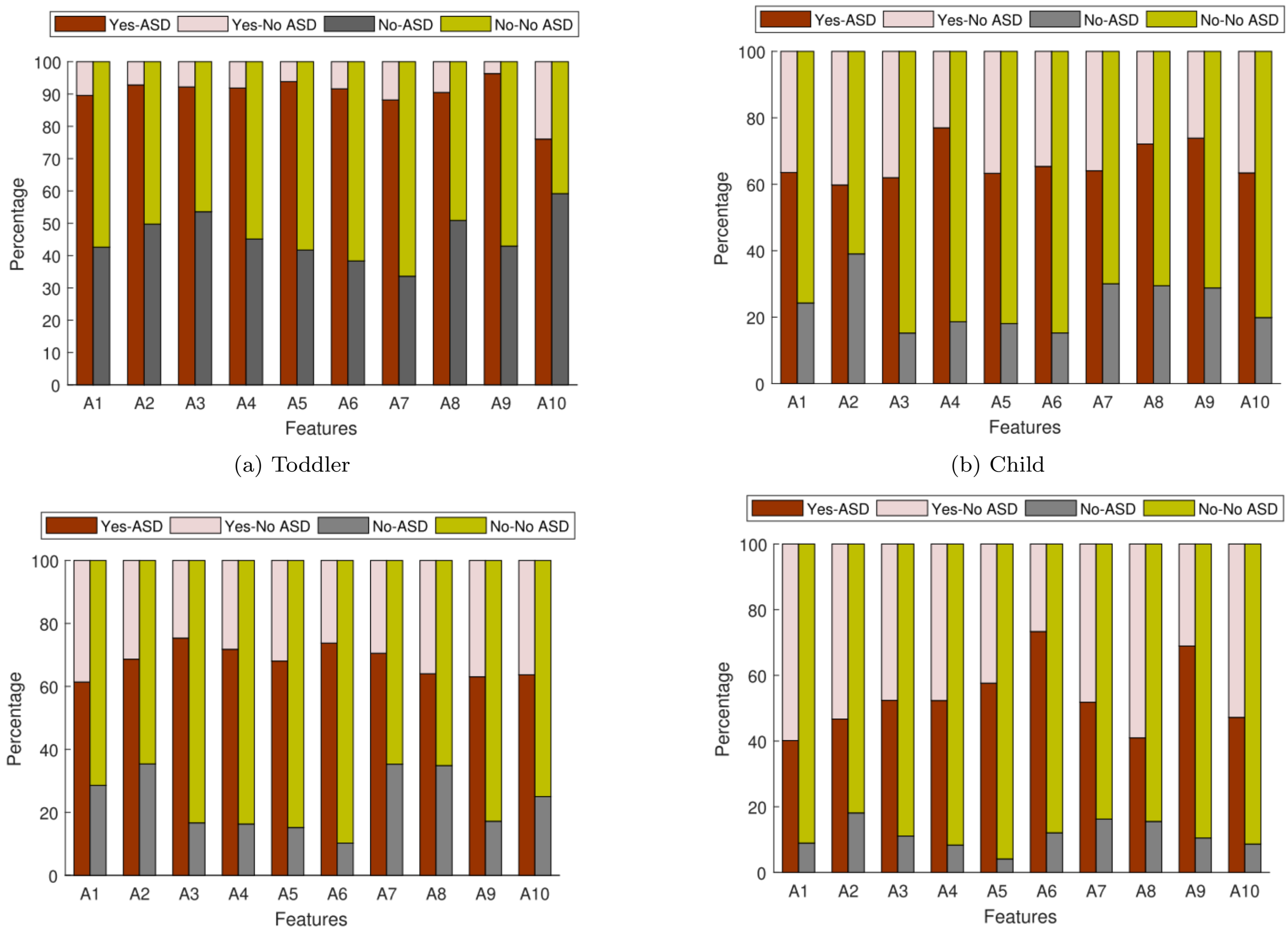
Attribute set	Size	MLP	SMO	LR	SL	LB	ICO	RAB	LMT
{A6}	1	79.44	79.44	79.44	79.44	79.44	79.44	79.44	79.44
{A6,A3}	2	81.85	81.85	81.85	81.85	81.85	81.85	81.86	81.86
{A6,A3,A5}	3	83.06	80.65	80.65	80.65	80.65	81.05	82.26	80.65
{A6,...,A4}	4	83.87	85.08	84.27	85.89	84.27	84.68	84.68	85.89
{A6,...,A7}	5	84.68	87.50	87.90	87.50	87.50	87.50	87.5	87.5
{A6,...,A9}	6	85.89	85.89	87.10	89.11	87.50	88.31	90.33	89.12
{A6,...,A10}	7	88.71	89.11	92.34	91.53	91.13	91.13	91.13	92.34
{A6,...,A2}	8	91.94	91.13	91.53	90.32	91.53	90.32	89.52	89.92
{A6,....,A1}	9	92.74	89.92	93.55	90.32	91.13	90.32	89.52	90.33
{A6,...,A8}	10	**100**	99.60	100	97.98	97.58	98.39	97.99	97.99
{A6,...,Ethn}	11	100	99.19	98.39	99.19	95.56	95.97	97.99	99.2
{A6,...,Res}	12	99.19	97.98	94.35	96.37	97.18	97.18	97.18	96.38
{A6,...,Age}	13	99.19	97.98	95.16	96.77	97.18	97.18	97.18	96.78
{A6,...,FASD}	14	99.19	97.98	95.56	96.37	97.18	97.18	97.18	96.38
{A6,...,Sex}	15	98.79	97.58	95.56	95.56	97.18	97.18	97.18	95.57
{A6,...,Jand}	16	98.79	97.18	94.76	95.97	97.18	97.18	97.18	95.97

Table 9: Classifiers' accuracy (%) with increasing number of attributes according to their ranks (from Relief F) for adult ASD dataset

Attribute set	Size	MLP	SMO	LR	SL	LB	ICO	RAB	LMT
{A5}	1	74.87	75.94	75.94	75.94	75.94	75.94	75.94	75.94
{A5,A6}	2	85.24	85.24	85.24	85.24	85.24	85.24	85.25	85.25
{A5,A6,A9}	3	87.30	86.49	87.75	87.66	87.75	87.66	87.39	87.66
{A5,...,A3}	4	90.70	90.70	90.70	90.70	90.70	90.70	90.7	90.7
{A5,...,A4}	5	89.98	90.34	90.16	89.62	90.16	90.16	90.26	89.54
{A5,...,A10}	6	91.14	91.68	91.32	90.61	90.25	90.43	91.42	90.61
{A5,...,A7}	7	92.84	94.01	94.01	94.01	94.10	94.01	94.01	94.01
{A5,...,A1}	8	94.36	95.80	95.44	95.62	94.72	94.81	95.62	95.62
{A5,...,A8}	9	95.35	95.71	96.15	96.06	95.89	95.71	95.89	96.07
{A5,...,A2}	10	**100**	**100**	**100**	99.82	99.91	99.91	99.83	99.83
{A5,...,Res}	11	100	100	98.39	99.82	99.91	99.91	99.83	99.83
{A5,...,Ethn}	12	100	100	97.85	99.82	99.91	99.91	99.83	99.83
{A5,...,Jand}	13	100	100	98.30	99.82	99.91	99.91	99.83	99.83
{A5,...,Age}	14	100	100	97.76	99.82	99.91	99.91	99.83	99.83
{A5,...,Sex}	15	100	100	97.41	99.82	99.91	99.91	99.83	99.83
{A5,...,FASD}	16	100	100	97.32	99.82	99.91	99.91	99.83	99.83

6.2 Identifying most significant features

We analyze the classifier performance with increasing number of attributes (as per their ranking) and compare the accuracy of eight selected classifiers. We observe that the accuracy increases with the increment of the number of attributes. It reaches to the maximum (for most of the classifiers) when the total number of attributes is 10, as shown in Table 6, 7, 8 and 9. After that the accuracy remains mostly constant for toddler, child



(a) Toddler (b) Child (c) Adolescent (d) Adult

Fig. 4: AQ-10 questions responses for ASD datasets

and adult but for adolescent it drops slightly with the increment of the number of attributes. Thus, we can argue that A1 to A10 attributes are the most significant attributes/features to accurately detect the ASD case. Although we see that MLP and Logistic Regression (LR) classifiers exhibit 100% accuracy for top ten attributes (which is the minimal number of attributes), MLP's accuracy remains same (i.e, 100%) for toddler, child and adult datasets with the increment of the attributes, whereas LR performance drops off after minimal attribute point. Moreover, for adolescent dataset LR performance drops is more than the MLP drops. Thus, we argue that MLP outperforms among all eight classifiers including LR.

7 Classification Results Comparison

In this section, we compare the prediction accuracy of this paper with state-of-the-art research. Most of the previous research works are based on on version-1 ASD dataset [19, 20, 22–25]. Among them only Baranwal et al. [23] have considered feature reduction while keeping the accuracy maximum.

To the best of our knowledge, Thabtah et al. [1] worked on version 2 adolescent and adult datasets only, and applied logistic regression classifier. The authors applied chi-squared and information gain feature ranking techniques to identify the most significant attributes and achieved 99% accuracy for adolescent dataset and 97.58% accuracy for adult dataset. Whereas, we have applied five feature ranking methods including chi-squared and information gain, and have found that Relief F feature ranking outperforms chi-squared and information gain by contributing to achieve 100% classification accuracy. Moreover, we have systematically select best classifier (from benchmark 27 classifiers), identify best performing feature ranking method (from five prominent methods), and determine minimal set of attributes

Table 10: Performance comparison between different ASD detection works

Paper	Dataset	Feature reduction	Best classifier	Toddler		Child		Adolescent		Adult	
				Recall	Accuracy	Recall	Accuracy	Recall	Accuracy	Recall	Accuracy
[19]	v1	✗	SVM	–	–	–	–	–	–	1	100
[20]	v1	✗	RF	–	–	–	–	–	–	0.93	91.74
[25]	v1	✗	RML	–	–	0.91	91	0.87	87.50	0.94	94.50
[22]	v1	✗	CNN	–	–	0.967	98.30	0.93	96.85	0.99	99.53
[23]	v1	✓	ANN, LR, SVM	–	–	1	96.77	1	80.0	0.98	98.90
[24]	v1	✗	RF	–	–	1	100	1	100	1	100
[1]	v2	✗	LR	–	–	–	–	0.99	99.91	0.97	97.58
This paper	v2	✓	MLP	1*	100*	1*	100*	1*	100*	1*	100*

Note: * result from top ten attributes

to achieve best accuracy. Furthermore, we have used all four version-2 ASD datasets.

Since our research is on version-2 dataset, we can only compare the work that used the same dataset such as Thabtah et al. [1]. In spite of this, we have added the version-1 related works in our comparison Table 10 to provide an overview of performance achieved in detecting ASD. The results presented in Table 10 show that this paper outperforms state-of-the-art research on ASD detection (irrespective of the dataset version), and the first work that used toddler ASD dataset.

8 Conclusion

In this study, we have analyzed the ASD datasets of toddler, child, adolescent and adult. We apply most popular five feature selection methods to derive fewer features from ASD datasets yet maintaining competitive performance. We find that Relief F feature selection method outperforms amongst others. In our experimental setup, we increase the attribute numbers gradually and then apply different classification techniques. We find that MLP outperforms amongst all other classifiers using our methodology and approach.

The main limitation of this research is the small datasets. In future work, we aim to collect large datasets and work with deep learning methods that integrate feature assessment and classification together for improved performance. Also,we would like to analyse brain signals (e.g., EEG) to relate this with AQ based study in order to develop a more robust ASD detection algorithm.

References

1. F. Thabtah, N. Abdelhamid, D. Peebles, A machine learning autism classification based on logistic regression analysis, Health Information Science and Systems **7**(1), 1 (2019)
2. L.D. Wiggins, A. Reynolds, C.E. Rice, E.J. Moody, P. Bernal, L. Blaskey, S.A. Rosenberg, L.C. Lee, S.E. Levy, Using Standardized Diagnostic Instruments to Classify Children with Autism in the Study to Explore Early Development, Journal of Autism and Developmental Disorders **45**(5), 1271 (2015)
3. C. Lord, M. Rutter, A. Le Couteur, Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders, Journal of Autism and Developmental Disorders **24**(5), 659 (1994)
4. A.L. Le Couteur, I. Gottesman, P. Bolton, E. Simonoff, E. Yuzda, M. Rutter, A. Bailey, Autism as a strongly genetic disorder evidence from a british twin Study, Psychological Medicine **25**(1), 63 (1995)
5. N.V. Chawla, in *Data Mining and Knowledge Discovery Handbook* (Springer US, Boston, MA, 2009), pp. 875–886
6. D. Geschwind, J. Sowinski, C. Lord, Letters to the Editor 463, The American Journal, 2001
7. A. Pratap, C.S. Kanimozhiselvi, R. Vijayakumar, K.V. Pramod, Soft Computing Models for the Predictive Grading of Childhood Autism-A Comparative Study. Tech. Rep. 4 (2014)
8. G. Fischbach, C.L. Neuron, The Simons Simplex Collection: a resource for identification of autism genetic risk factors, Elsevier (2010)
9. M.R. Islam, M.A. Kabir, A. Ahmed, A.R.M. Kamal, H. Wang, A. Ulhaq, Depression detection from social network data using machine learning techniques, Health information science and systems **6**(1), 8 (2018)
10. C. Allison, B. Auyeung, S. Baron-Cohen, Toward Brief "Red Flags" for Autism Screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1,000 Cases and 3,000 Controls. Tech. rep. (2012)
11. M. Duda, R. Ma, N. Haber, D.P. Wall, Use of machine learning for behavioral distinction of autism and ADHD, Translational Psychiatry **6**(2), e732 (2016)
12. F. Thabtah, Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment, dl.acm.org **Part F129311**, 1 (2017)
13. M.N. Parikh, H. Li, L. He, Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data, Frontiers in Computational Neuroscience **13** (2019)
14. F. Thabtah, *ASD Tests. A mobile app for ASD screening.* (accessed January 10, 2020). URL `https://www.asdtests.com/`
15. A. Zunino, P. Morerio, A. Cavallo, C. Ansuini, J. Podda, F. Battaglia, E. Veneselli, C. Becchio, V. Murino, in *24th International Conference on Pattern Recognition (ICPR)*, vol. 2018-August (IEEE, 2018), vol. 2018-August, pp. 3421–3426
16. S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, E. Clubley, The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning

Autism, Males and Females, Scientists and Mathematicians, Journal of Autism and Developmental Disorders **31**(1), 5 (2001)
17. F. Thabtah, *ASD Dataset- UCI machine learning repository, 2017* (accessed January 13, 2020). URL `https://archive.ics.uci.edu/ml`
18. F. Thabtah, *ASD Dataset* (accessed January 13, 2020). URL `https://fadifayez.com/autism-datasets/`
19. K. Basu, *Autism Detection in Adults Dataset.* URL `shorturl.at/ekyTW`
20. B. McNamara, C. Lora, D. Yang, F. Flores, P. Daly, *Machine Learning Classification of Adults with Autism Spectrum Disorder* (accessed June 25, 2020). URL `shorturl.at/hEFJY`
21. M.D. Hossain, M.A. Kabir, in *Proceedings of the 17th World Congress on Medical and Health Informatics (MedInfo 2019)*, vol. 264 (IOS, 2019), pp. 1447–1448
22. S. Raj, S. Masood, Analysis and detection of autism spectrum disorder using machine learning techniques, Procedia Computer Science **167**, 994 (2020)
23. A. Baranwal, M. Vanitha, Autistic spectrum disorder screening: Prediction with machine learning models, 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) pp. 1 – 7 (2020)
24. U. Erkan, D. Thanh, Autism spectrum disorder detection with machine learning methods, Current Psychiatry Reviews **15**, 297 (2019)
25. F. Thabtah, D. Peebles, A new machine learning model based on induction of rules for autism detection, Health informatics journal **26**(1), 264 (2020)