



# A new machine learning model based on induction of rules for autism detection

Health Informatics Journal

1–23

© The Author(s) 2019

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/1460458218824711

[journals.sagepub.com/home/jhi](https://journals.sagepub.com/home/jhi)**Fadi Thabtah**  and **David Peebles**

University of Huddersfield, UK

## Abstract

Autism spectrum disorder is a developmental disorder that describes certain challenges associated with communication (verbal and non-verbal), social skills, and repetitive behaviors. Typically, autism spectrum disorder is diagnosed in a clinical environment by licensed specialists using procedures which can be lengthy and cost-ineffective. Therefore, scholars in the medical, psychology, and applied behavioral science fields have in recent decades developed screening methods such as the Autism Spectrum Quotient and Modified Checklist for Autism in Toddlers for diagnosing autism and other pervasive developmental disorders. The accuracy and efficiency of these screening methods rely primarily on the experience and knowledge of the user, as well as the items designed in the screening method. One promising direction to improve the accuracy and efficiency of autism spectrum disorder detection is to build classification systems using intelligent technologies such as machine learning. Machine learning offers advanced techniques that construct automated classifiers that can be exploited by users and clinicians to significantly improve sensitivity, specificity, accuracy, and efficiency in diagnostic discovery. This article proposes a new machine learning method called Rules-Machine Learning that not only detects autistic traits of cases and controls but also offers users knowledge bases (rules) that can be utilized by domain experts in understanding the reasons behind the classification. Empirical results on three data sets related to children, adolescents, and adults show that Rules-Machine Learning offers classifiers with higher predictive accuracy, sensitivity, harmonic mean, and specificity than those of other machine learning approaches such as Boosting, Bagging, decision trees, and rule induction.

## Keywords

autism diagnosis, classification, decision-making, machine learning, predictive models, rule-based classifiers

## Introduction

Instances of autism spectrum disorder (ASD) are rapidly increasing. One in every 68 children is diagnosed with ASD, a developmental condition that presents with certain challenges associated with communication (verbal and non-verbal), social skills, and repetitive behaviors. It is estimated

---

### Corresponding author:

Fadi Thabtah, Digital Technologies, Manukau Institute of Technology, Corner of East Tamaki Road and, Newbury St, Otara, Auckland 2023, New Zealand.

Email: [Fadi.fayez@manukau.ac.nz](mailto:Fadi.fayez@manukau.ac.nz)

that 1.5 percent of the entire world population is classified with autism.<sup>1,2</sup> Unfortunately, the process of officially diagnosing individuals with autism is tedious, requiring clinical resources and diagnosis methods such as Autism Diagnostic Interview (ADI) and Autism Diagnostic Observation Schedule (ADOS).<sup>3,4</sup> Consequently, it is believed that many more people who are on the spectrum remain undetected.<sup>5</sup>

Moreover, the time spent waiting for a formal diagnosis is lengthy; for instance, the average waiting time in the United Kingdom is over 3 years.<sup>6</sup> Therefore, scholars in psychiatric health, psychology, and the behavioral science fields have developed self-administered and parent-administered screening methods that at a preliminary phase provide individuals with the recognition of possible autistic traits. Examples of screening methods are as follows: Screening Tool for Autism in Toddlers and Young Children (STAT), Childhood Autism Rating Scale (CARS-2), and Autism Spectrum Quotient (AQ).<sup>7–10</sup>

The accessibility and use of ASD screening tools are vital, as they may reduce waiting time for formal clinical evaluation and provide individuals on the spectrum, and their families, better understanding of the resources and services needed for support (special education, speech therapy, work environment, etc.). However, most existing screening tools are based on diagnostic methods that contain large numbers of items that the parent, caregiver, or the individual (in case of adult with an average intelligence quotient) is required to check. Therefore, these methods have been criticized as being too time-consuming.<sup>11–16</sup>

ASD traits are often screened using recognizable and measurable behavioral indicators (e.g. social skills, engagement in age-appropriate play and leisure, behavior excesses, and communication skills). These indicators are usually represented by items given in a questionnaire format for most current screening methods (i.e. AQ, STAT, CARS-2, etc.) The screening processes for individuals mainly rely on simple human rules with a scoring function that adds scores associated with the items in the questionnaire to calculate the outcome. Therefore, the quality of the classification outcome for individuals undergoing such screening is primarily based on (1) the items designed in the method, (2) the experience and knowledge of the user who is administering the screening, and more crucially (3) the handcrafted rules linked with the scoring function.

Designing rules to compute the scores of the questionnaire components requires extensive knowledge and experience. Replacing the human rule with knowledge derived from previous cases and controls to improve the diagnostic outcome and classification process seems advantageous. Automated knowledge is not subjective, as are handcrafted human rules, because they are discovered using advanced learning approaches such as machine learning (ML) or data mining, and consequently, boosting specificity, sensitivity, and predictive accuracy as well as classification efficiency. There is an urgent need for some advanced intelligent methods that can offer automatic classification of ASD as well as the reason(s) for the classification. These intelligent methods can be utilized by clinicians, parents, teachers, caregivers, and family members, among others, to understand the outcome of the screening. In addition, the clinician can use that outcome to verify the result of the screening using his or her own knowledge and experience.

Recently, a few scholars in the ASD research field have investigated ML to either improve the classification time of an ASD diagnosis or detect the most influential items in ASD diagnosis.<sup>12–14,17–21</sup> ML is a research area based around statistics, probability, artificial intelligence, databases, and other computer science areas that aims to intelligently discover hidden knowledge from data sets.<sup>22,23</sup> ML techniques, including support vector machines,<sup>24</sup> decision trees,<sup>25</sup> rule induction,<sup>26</sup> Boosting,<sup>27</sup> Bagging,<sup>28</sup> neural network,<sup>29</sup> and Covering,<sup>30</sup> seldom involve users in the processes of classification or model learning (Thabtah, 2017c).<sup>16</sup>

Since the ASD diagnostic process encompasses predicting whether individuals are on the spectrum, that is, with or without ASD, using predefined features including a class (ASD classification)

and a historical data set, this problem can be treated as a classification task in supervised learning. In this context, the clinician will utilize labeled cases of individuals with and without ASD (training data set) to construct a classification system (model) using an ML technique. The model is then employed to automatically forecast the class of a new case (cases that are not yet classified) as accurately as possible.

In this article, a new classification method based on the Covering approach, called Rules-Machine Learning (RML), is proposed. This method offers automatic classifications systems (classifiers) represented as rule sets. The rule sets inside the classifiers can be used by health professionals to assist in the diagnosis process or to advise individuals and their families whether they should seek further evaluation. The rules offered by the proposed method can be easily interpreted by novice users as well as parents, teachers, caregivers, and family members.

The RML was evaluated against real data sets collected using a mobile application called ASDTests<sup>31</sup> and recently published at the University of California Irvine Repository (UCI).<sup>32</sup> The experimental tests showed that the RML derives classifiers that are highly competitive when compared to other existing learning approaches in ML such as Boosting, Bagging, decision trees, and rule induction (section “Data features and empirical results” provides further details on the results and analysis). The performance evaluation of ML algorithms was based on common metrics such as predictive accuracy, sensitivity, harmonic mean, knowledge derived, and specificity.

This article is structured such that section “The problem and literature review” discusses the problem, aims, and critically analyzes related works on ML that are linked with ASD, while section “Proposed ASD classification system” presents the rule-based architecture and details related to the learning method. Finally, conclusions are presented in section “Data features and empirical results.”

## **The problem and literature review**

### *The problem and aims*

Official ASD diagnosis is typically conducted by specialist physicians in a clinical environment using a clinical judgment (CJ) procedure and based on observable and measurable behavioral indicators.<sup>21</sup> Existing paradigms seem to subscribe to the idea that more questions translate to a more accurate classification. Current standardized diagnostic tools take a very long time to conduct due to the large number of items that the specialist must check while relying on static human-embedded rules.<sup>33,34</sup> This has necessitated a change in the way diagnostics are coded and behave within ASD clinical tools in the process of classifying cases.

There is a need to re-examine features within ASD diagnostic tools to fulfill smaller item sets while maintaining the sensitivity and validity of the test. However, very limited examination of the ML perspective on autism has been previously conducted regarding the classification and validation processes of autism.<sup>15,20</sup> This new paradigm of utilizing ML will not only make pre-diagnostic tools faster and more accessible but will also dramatically change the perspective of designing clinical diagnostic tools. When the ML algorithm is embedded in self-assessment tools, it will provide users with valuable concealed knowledge and guide the process of correct classification selection decisions in a more efficient manner.<sup>14,35</sup>

To address this global dilemma, the proposed research article will take a new direction in the development of an ASD screening tool that incorporates rule-based architecture. Furthermore, this study aims to better understand what components contribute to an efficient and accessible data-based ASD screening tool such that may be used by health professionals and other stakeholders seeking to understand whether they should seek an autism diagnosis by a professional. More

specifically, we seek to establish a method that can be embedded within a self-administered ASD screening method to reliably and accurately provide feedback to patients, caregivers, and medical professionals regarding the potential need for professional diagnostic services. This investigation is vital for the standardization of efficient ASD diagnostic tools worldwide, serving to support long-term research goals and potentially impacting society directly.

This study also aims to limit the role of human-derived rules embedded within current assessment tools using ML technology to increase classification accuracy, sensitivity, and specificity. This is particularly necessary for cases that are difficult to classify (e.g. cases unclearly associated with an ASD type). Results obtained from the proposed ASD pre-diagnostic tool are expected to be initially utilized by medical professionals for more efficient referrals to comprehensive evaluations. The main research questions that this study will answer are as follows:

1. Is ML applicable to self-administered screening methods for ASD?
2. Can rule-based ML methods help ASD screening and diagnosis in terms of the efficiency, accuracy, and knowledge presented to the user?

### *Literature review*

There have been a few studies investigating ML in autism screening research in recent years (Thabtah, 2017c).<sup>12–15,17,33,36–38</sup> These studies have concentrated mainly on the following two aspects of ASD diagnosis:

1. Accelerating the diagnostic process by identifying the least number of items required to be checked during the screening;
2. Improving the sensitivity and specificity of the diagnostic decision by adopting an ML algorithm (neural network or decision tree) instead of the scoring function embedded in the diagnostic methods.

Wall and colleagues<sup>12,38</sup> conducted a comparative study using several ML algorithms, particularly decision tree-based, on a data set that contained cases and controls collected using the ADOS-Revised-Module 1 diagnostic method.<sup>4</sup> The data set was imbalanced with respect to class labels and contained many missing values and was stored in the Autism Genetic Resource Exchange (AGRE) repository.<sup>39</sup> The aim of the study was to calculate the most influential items in the ADOS-Revised-Module 1 that can be utilized by clinicians to reduce the time associated with the diagnosis. The authors utilized the WEKA platform to conduct the experiments of the different ML algorithms.<sup>40</sup>

Results obtained claimed that the ADOS-Revised can be replaced with merely eight items. The eight items were identified in classifiers generated by the Alternating Decision Tree Algorithm (ADTree) by simply looking at what items appeared in the classifiers. A more suitable approach is to investigate the impact of feature selection methods such as wrapping or filtering on the ASD data set and then analyze common features. A later study by Bone et al.<sup>13</sup> reported serious conceptual and implementation issues associated with the studies by Wall and colleagues.<sup>12,38</sup>

Duda et al.<sup>14</sup> reported conceptualization and implementation issues linked with the studies by Wall and colleagues.<sup>12,38</sup> The authors stressed that ASD prediction based on ML requires careful investigation especially when dealing with diagnostic methods that strictly follow procedures within a clinical environment. The claim that the ADOS-Revised diagnostic method can be minimized to eight items is misleading since to produce the decision, the entirety of activities must

be conducted by the clinician on a test case before the classification system is constructed. Consequently, there is no time saving.

More importantly, the activities of the ADOS-Revised must be performed in a clinical setup and not self-administered, as claimed by Wall and colleagues.<sup>12,38</sup> Therefore, the eight items proposed by Wall and colleagues, cannot replace the original items of ADOS-Revised. The full ADOS-R test must be conducted before building a classifier using the ADTree algorithm in WEKA. Finally, discarding, by the authors, data that are on the border between ASD and No ASD simplified the problem. This is because these cases are hard to detect by the ML algorithms. Therefore, including them prior to the data processing phase will impact the sensitivity, specificity, and accuracy of the results.

Duda et al.<sup>14</sup> conducted an empirical analysis comparing several intelligent algorithms to discriminate between ASD and attention deficit hyperactivity disorder (ADHD). Six algorithms have been contrasted on a data set with 65 items adopted from the Simons Simplex Collection (SSC) version 15.<sup>41</sup> The data set was collected using a parent-administered questionnaire diagnostic method called Social Responsiveness Scale (SRS).<sup>42</sup> A pre-processing phase was applied by the authors to (1) discard instances that had four or more missing values, (2) balance the data set using under sampling technique, and (3) reduce the data dimensionality using feature selection methods. Empirical results reported that logistic regression produced classifiers with almost 95 percent classification accuracy.

Chu et al.<sup>43</sup> investigated efficient ways to differentiate between ADHD and obstructive sleep apnea (OSA). The authors utilized the information of 217 children who had been classified by physicians as having ADHD, OSA, and a mixture of ADHD and OSA according to *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; DSM IV) standards.<sup>44</sup> The data were collected using different diagnostic tools. Three ML algorithms were adopted to derive classifiers that could assist clinicians and physicians in improving the diagnostic decision. Reported results indicated that 17 features show substantial difference among three classes of pervasive developmental disorders (PDDs) particularly in the Child Behavior Checklist (CBCL).<sup>45</sup> A decision tree algorithm called CART was faster to derive the classifiers than the neural network and CHAID algorithms.

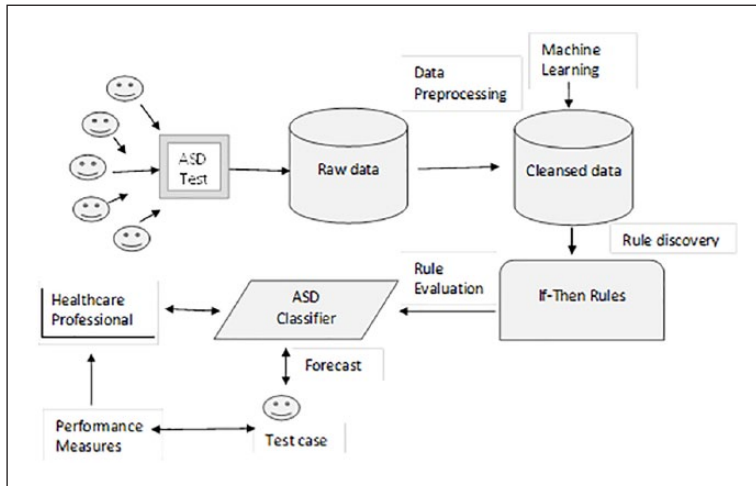
Wolters et al.<sup>37</sup> investigated issues related to PDDs including small sample sizes, external validity, and ML algorithmic challenges without a clear focus on ASD. Lopez Marcano<sup>33</sup> reviewed the applicability of different algorithms such as neural network and decision tree methods (Random Forest) to shorten the time of the ASD diagnostic process. Maenner et al.<sup>46</sup> investigated the Random Forest algorithm<sup>47</sup> on an autism data set from Georgia Autism and Developmental Disabilities Monitoring (ADDM) Network utilizing phrases and words obtained in children's developmental evaluations. The data set consists of 5396 evaluations for 1162 children of whom 601 are on the spectrum. The Random Forest classifiers were evaluated on an independent test data set that contained 9811 evaluations of 1450 children. The results reported that Random Forest achieved around 89 percent predictive value and 84 percent sensitivity.

Thabtah (2017c) critically analyzed pitfalls associated with experimental studies that adopted ML for ASD classification. The authors pinpointed issues related to data sets and learning algorithm methodologies used. These issues included interpreting the classifiers content derived by the learning algorithm, noise in autism data sets, feature selection process, missing values, class imbalance, and embedding the classification algorithm within an existing screening method.

## Proposed ASD classification system

### *Rule-based architecture for detecting ASD (RML)*

One of the least studied classification approaches in ML is Covering. Covering techniques normally discover simple chunks of information from historical data sets structured in the If-Then



**Figure 1.** The proposed ML architecture for ASD classification.

format, which makes their outcome highly favorable to novice users. In this section, we propose a new ASD detection method based on the architecture shown in Figure 1. Our method (RML) is based on Covering classification which employs a search method for rule discovery. The RML then evaluates the discovered rule and discards any redundancies. Hence, only rules that have been classified as training instances are kept. The evaluation phase performed not only reduces the number of discovered rules but also shrinks the search space of data items, which improves the efficiency of the training process.

In Figure 1, data are collected by a mobile application called ASDTests<sup>31</sup> that implements four different ASD screening methods (AQ-Adult-10, AQ-Adolescent-10, AQ-Child-10, and Q-CHAT-10).<sup>11</sup> For the purposes of this research project, focus was on the child, adolescent, and adult modules and researchers utilized three data sets collected between September 2017 and January 2018. Once the raw data were obtained, several pre-processing operations were applied, including missing values replacement and discretization for certain continuous attributes such as the age of individuals.

Feature selection was used to remove features that were redundant and may have created biased results. Two features were eliminated, including the final score obtained by the screening method and the scoring method type (see section “Data features and empirical results” for further details on data features).

Once the raw data are preprocessed, then a learning algorithm is applied to discover rule sets that represent correlations between the variables in the training data set and the class variables (ASD or No ASD). The data sets are then evaluated to remove useless and redundant rules, storing only rules that have classified training instances.

The outcome of the rule evaluation phase is the classification system (classifier) that will be used to predict the value of the class for unseen cases (individuals who have not yet been classified). When the classifier is tested, various evaluation metrics are derived to reveal the effectiveness of the rules in predicting cases and controls. These metrics, as well as the rules in the classifier, are shared with the health professional and the users involved in the screening. Therefore, not only does the new architecture provide users with decisions related to ASD detection, it also offers rich information on the reasons behind that decision as well as the quality of the outcome.



**Algorithm 1.** The RML classification method.

Input: data set with cases and controls  $T$ ,  $R\_S$ ,  $Min\_Freq$  thresholds // Rule Strength =  $R\_S$ . Minimum Frequency =  $Min\_Freq$ .

Output: A Model with rules ( $RS$ )

1.  $E\_S\_R \leftarrow \{\}$
2.  $r_i \leftarrow \{\}$
3.  $Temp \leftarrow T$
4. Do {
5. If  $[(p(A_i, v_i)|c_i=l)/|Temp|] \geq min\_freq$  {
6. If  $[(p(A_i, v_i)|c_i=l)/|p(A_i, v_i)|] \geq R\_S$  {
7.  $r_i \leftarrow (A_i, v_i)$
8. Repeat steps 5–7 until  $r_i$  accuracy cannot improve  
    }}
9.  $E\_S\_R \leftarrow r_i$
10.  $Temp \leftarrow (Temp - Locate(r_i, Temp))$
11.  $Train \leftarrow (Train - Temp)$
12. Repeat steps 2–11
13. Exit when  $T$  has no more instances OR all  $p(A_i, v_i)$  have been tested
14. }
15. Generate  $E\_S\_R$
16. Classify Test ( $Test, E\_S\_R$ )

**The learning Covering algorithm**

Researchers here developed a new learning mechanism based on the Covering approach called RML. The learning method pseudocode is shown in Algorithm 1. The algorithm utilizes two thresholds named the Minimum Frequency ( $Min\_Freq$ ) and Rule Strength ( $R\_S$ ) as other Covering approaches such as Dynamic Rule Induction (DRI)<sup>30,48</sup> to find and extract the rules (Definitions 2 and 3, respectively). The  $Min\_Freq$  threshold is used as a cutoff point for variables and class values in the training data (items).

An item is represented as (Variable Value, Class Value) (Definition 1), and any item in the training data with a frequency equal to or above the  $Min\_Freq$  threshold is qualified to be part of the rule's body during the process of building a rule. On the other hand, each rule is linked with a calculated strength (Rule Strength), which denotes the rule's items plus class frequency divided by the items' frequency (see Definition 5).

A rule is represented as  $(A_1, v_1) \wedge (A_2, v_2) \wedge \dots \wedge (A_k, v_k) \rightarrow C_n$  where the antecedent is a conjunction of variables values (rule body), and the consequent is a class value (ASD, No ASD). When the computed rule strength for a rule such as  $R$  is larger than or equal to the  $R\_S$  threshold,  $R$  can then be generated, otherwise  $R$  will be removed. The computed rule strength for any given rule acts as a quality assurance metric that ensures only mathematically fit rules (that have proper data representation) are generated.

The  $Min\_Freq$  and  $R\_S$  thresholds are like minimum support and minimum confidence parameters in association rule mining.<sup>49</sup> However, minimum support and minimum confidence are used to differentiate frequent items from infrequent items by considering the items' frequencies in the transactional data, whereas  $Min\_Freq$  and  $R\_S$  thresholds consider the target class when counting attribute values. More importantly, whenever the rule is derived by RML, the data set is amended and therefore the frequency of rules is updated.

**Definition 1.** 1-Item in the training data set ( $T$ ), that is,  $[(A_1, v_1), C_n]$ , is an attribute plus a class. K-Item is a combination of attributes values plus a class, that is,  $[(A_1, v_1), (A_2, v_2), \dots, (A_k, v_k), C_n]$ .

**Definition 2.** Min\_Freq is a user threshold used to separate weak items from strong items.

**Definition 3.** R\_S is a user threshold used to form rules.

**Definition 4.** A strong item, that is,  $[(A_1, v_1), C_n]$ , is recognized when  $|[(A_1, v_1), C_n]| / T \geq \text{Min\_Freq}$ .

**Definition 5.** A rule such as  $r$  is formed when  $|[(A_1, v_1), (A_2, v_2), \dots, (A_k, v_k)], C_n]| / |[(A_1, v_1), (A_2, v_2), \dots, (A_k, v_k)]| \geq R\_S$ .

The learning algorithm initially scans the training data set and discards any 1-item that has failed the Min\_Freq threshold test (lines 4–5). All remaining items with computed frequencies above the Min\_Freq threshold are considered and saved into a data structure. To build a rule such as  $R$ , the algorithm attaches the best item in terms of computed frequency to the rule's body and repeats the process until the rule's accuracy cannot improve any further (lines 6–7). When this occurs, the rule is then saved into the classifier (line 9) and all training instances linked with  $R$  are erased from the original training data set (lines 10–11). When this happens, the strong items' frequencies are updated in the data structure. Consequently, some items may become weak and thus discarded by the learning algorithm.

In other words, items that share training instances with  $R$  are affected by data removal; therefore, frequencies of these items are normally reduced. The update procedure ensures that rules learnt are indeed non-redundant and often cover a larger portion of the training data set. Continuing, this procedure can be considered as a quality measure, as items' frequencies are continuously updated since the training data set is shrinking whenever a rule is generated. Resulting from this repetitive learning process, some manageable models with small yet effective rules are formed, which then can be exploited for decision-making by users in the autism screening application.

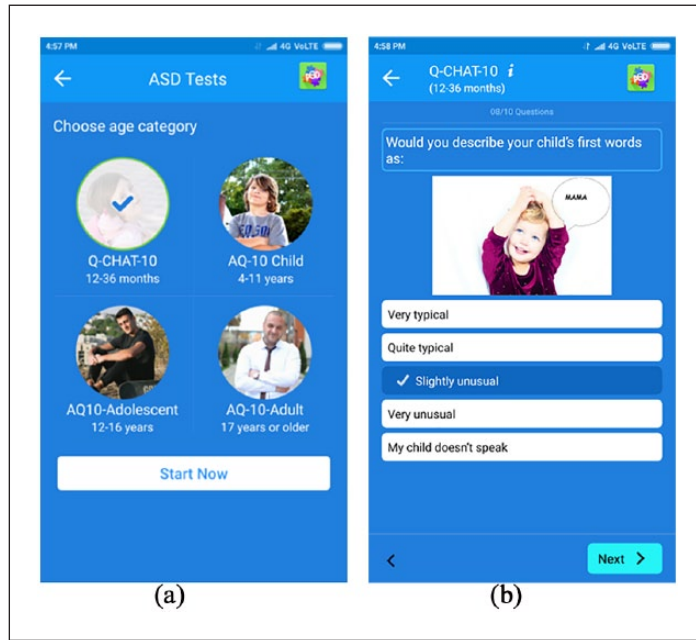
The RML algorithm guarantees that the search space of items is constantly reduced during the training phase and thus results in more efficient data processing. In addition, data instances that might overlap among items are removed, ensuring that rules extracted are not similar. Recall that RML keeps appending items in the rule's body until it processes with zero error so that the rule can be derived. However, in scenarios when rules are associated with some errors, RML allows the generation of such rules if they have computed strengths larger than or equal to the  $R\_S$  threshold set by the end user. This mechanism offers rules with slightly acceptable margins of error but minimizes the chance of models' overfitting.

The RML assumes that the variables in the training data set are categorical (they are associated with a finite set of possible values). Continuous variables (integers and decimals) should be discretized before data processing. Finally, missing values are dealt with as any other values in the training data set.

To evaluate the rule sets generated by the learning algorithm, a test procedure that assigns test data the appropriate class is utilized. Whenever a test case is present, the test method allocates the class label linked with the best ranked rule that matches the test case. This method necessitates that all items of the selected rule's body are presented in the test case for the rule to be used for prediction. In cases when there are no rules in the classifier fulfilling this condition, the test method then allocates the class label of the first partially matching rule to the test case. When no rules are partially or fully matching the test case, then a default class is allocated. The default class is basically a rule that represents the class with the largest frequency in the training data set.

Hereunder are the key features of the ASD rule-based model:





**Figure 2.** (a) Age selection screen and (b) a sample question: toddler.<sup>31</sup>

1. The learning method produces non-redundant rules in the format “If–then” that are easy to understand by different users such as clinicians, physicians, family members, caregivers, teachers, and others;
2. Efficient procedure for learning the rules that requires one data scan and keeps reducing the search space of items during the training process;
3. Straightforward metrics are utilized to derive the rules;
4. Classifiers derived have fewer rules which make them more manageable by the different users;
5. Better sensitivity, specificity, and classification accuracy than the classical process-based scoring functions in current screening methods (see section “Data features and empirical results” for further details on the results).

## Data features and empirical results

### Data description

The data have been collected using a recently developed mobile application called ASDTests (Thabtah, 2018).<sup>50</sup> ASDTests implements four screening methods for toddlers, children, adolescents, and adults based on Q-CHAT-10, AQ-10-child, AQ-10-Adolescent, and AQ-10 adult, respectively.<sup>11</sup> Figure 2(a) and (b) depicts the landing page and one question page related to toddler test of the ASDTests app. During the data collection, there was no direct access to participants; the ASDTests mobile application provides clear information to the users about their participation and the use of their data in a disclaimer. In addition, the webpage also clearly states that the use of the data is for research purposes only and informs the users about the use of data. The participants read this before submitting their answers. Anonymity has been imposed in the mobile app used to collect the

**Table 1.** Features collected, their descriptions, and mapping to the actual AQ-10 questionnaire.

Feature	Type	Description
Age	Number	Adults (year), that is, 17 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	String	Parent, self, caregiver, medical staff, clinician, and so on
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 chosen based on age category (0= Toddler, 1=Child, 2=Adolescent, 3=Adult). In our case, only Adult data have been used
A1	Binary (0, 1)	The answer code of: I often notice small sounds when others do not
A2	Binary (0, 1)	The answer code of: I usually concentrate more on the whole picture rather than the small details
A3	Binary (0, 1)	The answer code of: I find it easy to do more than one thing at once
A4	Binary (0, 1)	If there is an interruption, I can switch back to what I was doing very quickly
A5	Binary (0, 1)	The answer code of: I find it easy to “read between the lines” when someone is talking to me
A6	Binary (0, 1)	The answer code of: I know how to tell if someone listening to me is getting bored
A7	Binary (0, 1)	When I’m reading a story I find it difficult to work out the characters’ intentions
A8	Binary (0, 1)	I like to collect information about categories of things (e.g. types of car, types of bird, types of train, and types of plant)
A9	Binary (0, 1)	I find it easy to work out what someone is thinking or feeling just by looking at their face
A10	Binary (0, 1)	The answer code of: I find it difficult to work out people’s intentions
ASD score	Integer	The final score obtained based on the scoring function of on AQ-10-Adult. This was computed in an automated manner
Class label	Boolean	The decision of the screening based on the scoring score of AQ-10-Adult method. Possible values “0” (no ASD traits or “1” (ASD traits))

PDD: pervasive developmental disorder; ASD: autism spectrum disorder.

data. Participants’ identities are anonymous since no names or sensitive information are involved (see variables in Table 1).

While using the ASDtests application, the user can choose the age category test, which includes 10 questions presented in a simple graphical user interface. Each question is associated with

multiple answers that are easily selected in a mobile environment using a smartphone (iOS and Android) or a tablet. Once all the questions have been answered by the user, a review screen appears so the user can review and verify their answers before a data submission page with a disclaimer is invoked. The data sets used have been recently published by the authors at the University of California Irvine Data Repository.<sup>32</sup>

Table 1 shows the key features in the data set used plus the type of screening, that is, class label. The class label was assigned in an automatic manner during the process of data collection by the AQ-10 scoring method based on the final score obtained by the individual after taking the screening. There are two possible values of the class, that is, “0” indicates that the individual has no ASD traits, and “1” indicates that the individual does have ASD traits. The “0” label is assigned when the final score based on the AQ-10 methods’ scoring function is more than 6. More details on the score calculation can be found in the work by Allison et al.<sup>11</sup> Overall, there are 20 features in Table 1 including the class label.

The 10 questions, A1–A10, have been transformed into binary attributes based on the values assigned to them by individuals during the screening process. To be exact, values in the A1–A10 variables in the data set have been mapped to “0” or “1” depending on the actual values given during the screening process by the participants. In other words, during the screening using the AQ-10 screening method, “1” was given for “Definitely” or “Slightly Agree” answers for questions 1, 7, 8, and 10. For the rest of the questions in this method, “1” was allocated when “Definitely” or “Slightly Disagree” was chosen for questions 2, 3, 4, 5, 6, or 9. The binary representation for features allows more efficient data processing when adopting learning algorithms in addition to making interpretation easier.

Table 2 displays 10 sample cases of individuals who experienced the AQ-10 adult screening, for presentation purposes. The data set size is 704 cases collected over a period of 4 months. The data set is imbalanced with respect to class labels, with 515 cases belonging to “No ASD Traits” and 189 cases with “ASD.” This can be attributed to the fact that most people being screened through the app have no autistic traits. There are missing values in some cases, especially in two features, that is, ethnicity, and `who_is_taking_the_test`. Slightly more male than female individuals have taken the ASD screening using the app. The three most popular ethnicities in the data set belong to White, Asian, and Middle Eastern. The computed mean for age was 29.2 with the youngest person to have taken the screening being 17 years old and the oldest 64 years old. Finally, more adults in the data set have taken the test independently. The adolescent and child data sets contain 104 and 292 instances, respectively.

### *Experimental settings*

This section presents the experimental settings of the proposed rule-based model and other common ML algorithms based on rule induction, Bagging, Boosting, and decision tree approaches on the child, adolescent, and adult data sets. We used six different algorithms in addition to RML to reveal the performance of the rule-based model. RIPPER, RIDOR, Nnge, Bagging, CART, C4.5, and PRISM algorithms have been adopted in the experimental results.<sup>26,28,51–55</sup> The main reason for choosing these algorithms, aside from them all producing rule-based classification models (classifiers), is the fact that they employ different learning schemes in processing the data set.

C4.5 and CART construct decision tree classifiers that get converted into rule sets; PRISM is a greedy algorithm that seeks for rules that have 100 percent expected accuracy. C4.5 uses pessimistic error estimation for pruning the trees before converting these trees into rule sets, whereas PRISM uses expected accuracy to measure the usefulness of adding an item into the rule’s body while constructing a rule. On the other hand, RIPPER and RIDOR implement optimization and pruning procedures to test rules. For instance, RIPPER uses growing and pruning data sets to

**Table 2.** Sample of 10 data cases collected for adults using ASDTests app based on AQ-10 adult screening method.

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Age	Sex	Ethnicity	Jaundice	Family with autism	Country_of_res.	Used_app_before	ASD score	Who is taking the test	Class/ASD
1	1	1	1	0	0	1	1	0	0	26	F	White	No	No	USA	No	6	Self	No
1	1	0	1	0	0	1	1	0	1	24	M	Latino	No	Yes	Brazil	No	5	Self	No
1	1	0	1	1	0	1	1	1	1	27	M	Latino	Yes	Yes	Spain	No	8	Parent	Yes
1	1	0	1	0	0	1	1	0	1	35	F	White	No	Yes	USA	No	6	Self	No
1	0	0	0	0	0	1	1	0	0	40	F	Arab	No	No	Palestine	No	2	Self	No
1	1	1	1	1	0	1	1	1	1	36	M	Others	Yes	No	USA	No	9	Self	Yes
0	1	0	0	0	0	1	1	0	0	17	F	Black	No	No	USA	No	2	Parent	No
1	1	1	1	0	0	0	0	1	0	64	M	White	No	No	New Zealand	No	5	Self	No
1	1	0	0	1	0	1	1	1	1	29	M	White	No	No	UK	No	6	Self	No
1	1	1	1	0	1	1	1	1	0	17	M	Asian	Yes	Yes	China	No	8	Healthcare professional	Yes

**Table 3.** Confusion matrix for ASD screening problem.

Actual class value	Predicted class value	
	ASD	No ASD
ASD	True positive (TP)	False negative (FN)
No ASD	False positive (FP)	True negative (TN)

ASD: autism spectrum disorder.

evaluate the worthiness of attributes' values prior to appending them into the rule's body. So, if adding an attribute value decreases the rule's predictive power, RIPPER ignores adding the attribute value and generates the rule. Finally, Bagging and Boosting employ weak classifiers that in turn are merged to derive rules. This has been accomplished by deriving  $N$  classifiers and then utilizing them in predicting the class label of test instances using a voting mechanism, that is, the class that belongs to the majority classifiers gets allocated to the test instance.

The considered algorithms are well investigated on different real-world applications and have proved their merits in terms of performance, such as predictive power and efficiency. Different evaluation metrics have been adopted to reveal the ML algorithm's true performance in detecting ASD traits from the data sets. To be exact, classification accuracy, specificity, and sensitivity, among others (see equations (1) to (5)) were used.<sup>56</sup>

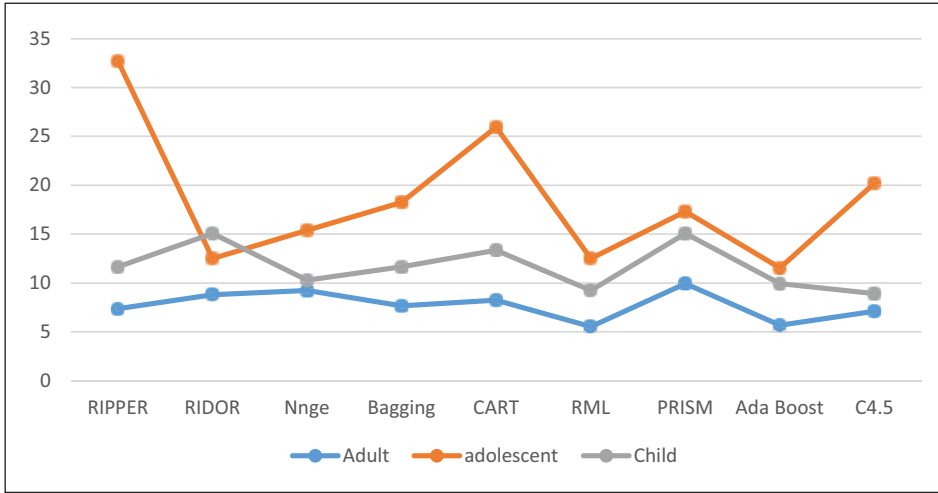
The proposed rule-based model has been developed in the Java programming language and integrated within the WEKA platform version 3.9.1.<sup>40</sup> WEKA is a known environment for implementing methods related to learning, classification, prediction, variable analysis, visualization, and dimensionality reduction. WEKA consists of packages that contain large numbers of ML and data mining techniques. Hence, all empirical runs have been conducted in WEKA for fair comparison. In testing the classifiers generated by the learning algorithm considered, a 10-fold cross-validation method has been adopted.<sup>56,57</sup>

In using 10-fold cross-validation, the training data set is partitioned into 10 subsets. The classification algorithm randomly utilizes nine data subsets to learn the classifier and then tests the classifier on the remaining data subset. The same process is repeated 10 times to generate an average error rate.<sup>50</sup> The cross-validation procedure is embedded in WEKA platform and can be selected prior the learning phase. Finally, all experimental runs have been conducted on a personal computer that has a 2.0GHz processor and 8RAM of memory.

The ASD screening process is a binary classification problem since individuals are classified to either having ASD traits or No ASD traits using characterized quantifiable variables. Therefore, performance evaluation methods that align with the binary classification problem in ML have been used. The confusion matrix (Table 3) can be used to derive different evaluation metrics including classification accuracy, error rate, sensitivity, and specificity to report the performance of the learning algorithms. Using the confusion matrix, a test case will be assigned a predicted class in the classification step of the screening.

Based on Table 3, classification accuracy (equation (2)) is a common metric in classification that computes the number of test data that were correctly classified from the total number of test data. Opposite to accuracy is the error rate (equation (1)). On the other hand, sensitivity (equation (3)) computes the percentage of the test cases that are truly positive (with ASD class) and specificity (equation (4)) denotes the percentage of the test cases that are negative (cases with no ASD)

$$One\_error(\%) = 1 - Accuracy \quad (1)$$



**Figure 3.** Error rates derived by the considered ML algorithms on the child, adolescent, and adult data sets.

$$Accuracy(\%) = \frac{|TP + TN|}{|TP + TN + FP + FN|} \quad (2)$$

$$Sensitivity(\%) = \frac{|TP|}{|TP + FN|} \quad (3)$$

$$Specificity(\%) = \frac{|TN|}{|TN + FP|} \quad (4)$$

$$Precision = \frac{|TP|}{|TP + FP|} \quad (5)$$

$$F1 = 2 \times \frac{|Precision \times Recall|}{|Precision + Recall|} \quad (6)$$

### Results analysis

Multiple experimental runs have been performed using different ML algorithms on child, adolescent, and adult ASD screening data sets to reveal the true performance of the proposed model. Figure 3 depicts the error rate results in percentage of the considered algorithms on the child, adolescent, and adult data sets. The figures show that Bagging, Boosting, rule induction, and decision



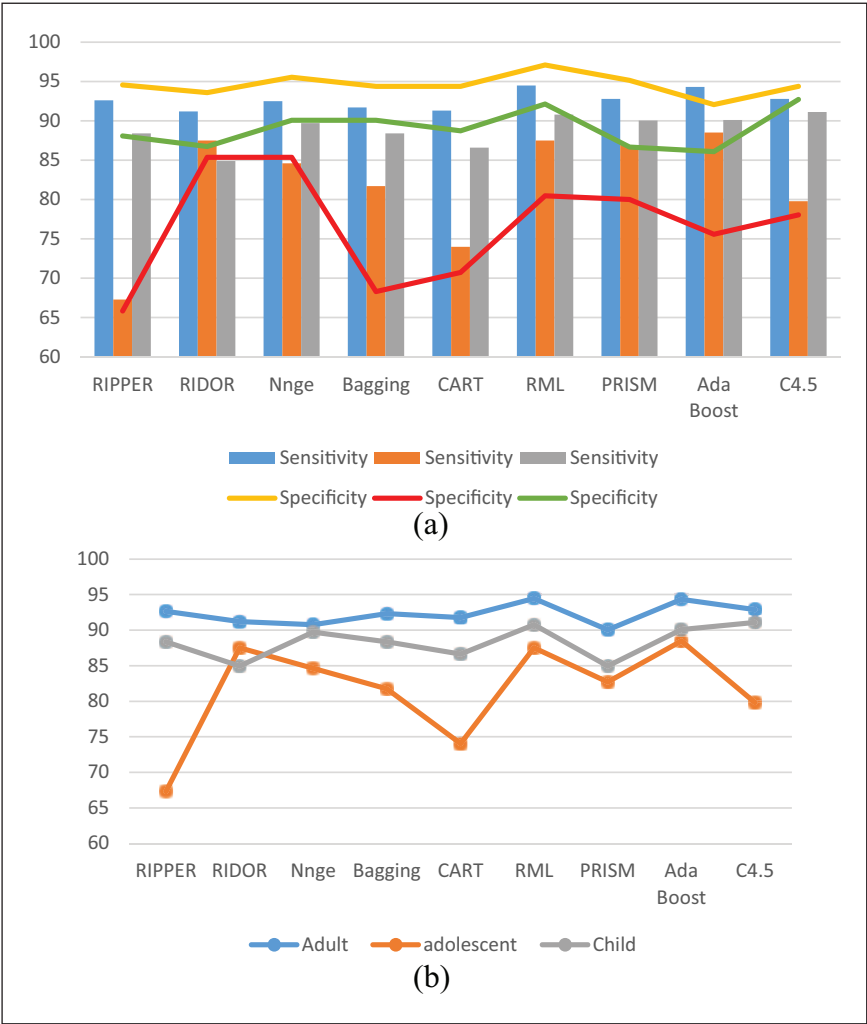
tree classifiers were able to accurately classify most of the cases and controls as their error rates for the adult data set were between 5.68 and 8.23 percent. However, the enhanced Covering algorithms such as our model (RML) outperformed the remaining algorithm in terms of error rate, that is, an error rate less than 5.6 percent.

For the adult data set, RML derived a classifier with lower errors rates of 4.41, 2.7, 0.15, 2.14, 3.7, 3.27, 1.57, and 1.83 percent, respectively, than PRISM, CART, AdaBoost, Bagging, Nnge, RIDOR, C4.5, and RIPPER algorithms. For the smaller data sets (adolescent, child), RML maintained higher predictive rates than most of the considered algorithms. For instance, for the child data set, RML achieved 5.82, 4.11, 0.69, 2.4, 1.03, 5.82, and 2.4 percent less error rate than PRISM, CART, AdaBoost, Bagging, Nnge, RIDOR, and RIPPER algorithms. Only C4.5 slightly achieved 0.34 percent higher than RML on this data set. Nevertheless, RML outperformed C4.5 on the adolescent data set by 7.69 percent. This, if limited, shows that RML not only performs well on data sets with enough data instances, such as the adult data set, but also with data sets with a limited number of instances, such as the adolescent data set. In addition, the superiority of the proposed algorithm is clear in the case of small data sets, whereas the considered ML algorithms suffered from low predictive rates due to not having enough instances. For example, rule induction algorithms, such as RIPPER, and tree-based algorithms, such as C4.5 and CART, derived classifiers with 20, 7.69, and 13.46 percent less error rates, respectively, than that of the RML model on the adolescent data set.

The reduction in the error rate can be attributed to the procedure employed by RML in the rule generation phase in which only non-redundant rules are produced and redundant rules that have no data coverage are discarded. Our model ensures that each rule has data coverage and eliminates any overlapping among rules on training instances, hence deriving an accurate classifier. In building the classification systems for detecting ASD, the RML algorithm ensures that whenever a rule is generated, all its data instances are removed before learning the next rule from the training data set. In addition, it amends candidate item frequencies during the learning phase whenever training instances associated with the generated rules are erased. These amendments may result in potential rules becoming weak and thus discarded at the preliminary phase, which reduces the search space and improves the efficiency of the training phase.

Figure 4(a) displays the specificity and sensitivity rates derived by the RIPPER, RIDOR, Nnge, Bagging, AdaBoost, CART, RML, C4.5, and PRISM algorithms on the child, adolescent, and adult data sets. Usually, acceptable specificity and sensitivity rates in autism research should be at least 80 percent.<sup>1</sup> The results of the specificity and sensitivity rates generated by the considered algorithms on the two data sets (adult and child) have shown acceptable levels. Moreover, the Covering approach represented by RML produced classification systems with higher sensitivity and specificity rates than most of the remaining algorithms on these data sets. For example, for the adult data set, RML derived 1.9, 3.3, 2.0, 2.8, 3.2, 1.7, 0.2, and 1.7 percent higher sensitivity rates than RIPPER, RIDOR, Nnge, Bagging, CART, PRISM, AdaBoost, and C4.5 algorithms, respectively. On the other hand, and for the same data set, RML achieved 2.52, 3.49, 1.55, 2.72, 2.72, 1.94, 5.02, and 2.72 percent higher specificity rates than RIPPER, RIDOR, Nnge, Bagging, CART, PRISM, AdaBoost, and C4.5 algorithms, respectively.

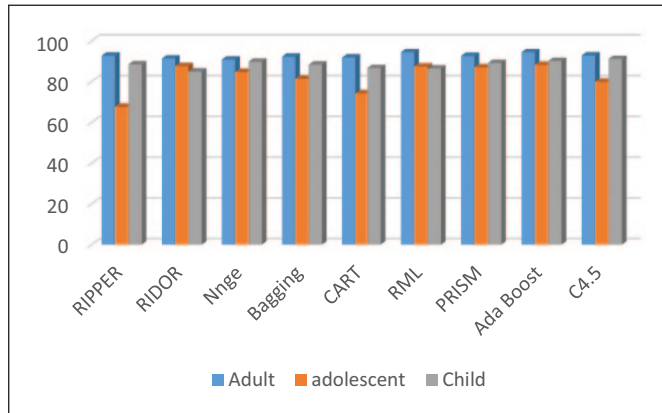
For the child data set, RML achieved 2.4, 5.9, 1.1, 2.4, 4.2, 0.78, and 0.7 percent higher sensitivity rates than RIPPER, RIDOR, Nnge, Bagging, CART, PRISM, and AdaBoost algorithms, respectively. The rates increase for the adolescent data set since most of the ML algorithms are unable to perform well on small data sets with lesser number of instances as RML. To be exact, the sensitivity rate of RML is 20.2, 2.9, 5.8, 13.5, 0.6, and 7.7 percent higher than RIPPER, Nnge, Bagging, CART, AdaBoost, and C4.5 algorithms, respectively. C4.5 slightly outperformed RML with respect



**Figure 4.** (a) Specificity and sensitivity rates of the ML algorithms on the adult, adolescent, and child data sets and (b) predictive accuracy rates in percentage derived by the considered ML algorithms on the adult, adolescent, and child data sets.

to sensitivity rate on the child data set and by 0.3 percent. For the child data set, the specificity rate of RML was higher than most of the considered ML algorithms. To be exact, RML derived 19.8, 2.7, 6.0, 13.2, 0.4, and 7.5 percent higher specificity rate than RIPPER, Nnge, Bagging, CART, PRISM, and C4.5 algorithms, respectively. Only RIDOR and AdaBoost slightly outperform RML in terms of specificity rate on the adolescent data set and by just 0.2 and 0.8 percent, respectively. Overall, the results reported higher sensitivity and specificity rates for RML on the three data sets when compared with the considered ML algorithms; these results are consistent with the error rates produced earlier and can be attributed to the non-redundant rule sets generated by RML.

The researchers investigated the confusion matrix results produced by the classifiers to understand the sensitivity, accuracy, and specificity results. For the adult data set, it was observed that the PRISM algorithm had the largest number of false negatives, followed by the CART and



**Figure 5.** F1 rates in percentage derived by the considered ML algorithms on the adult, adolescent, and child data sets.

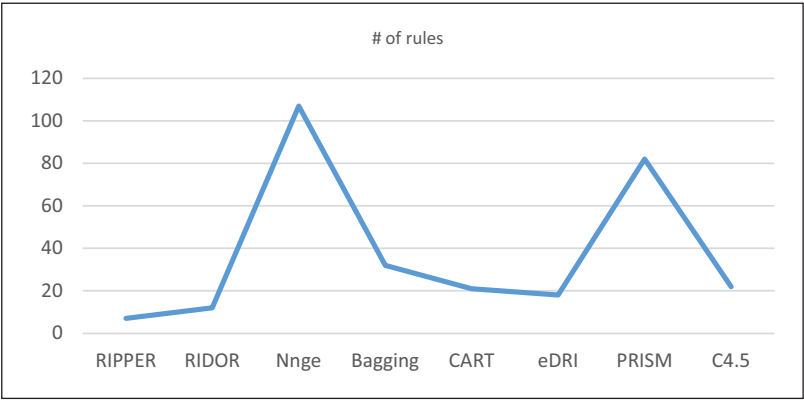
Bagging algorithms. Specifically, PRISM predicted 38 instances of individuals with No ASD traits that should have been classified on the spectrum. As a result, the sensitivity rate for class “ASD” for this algorithm was low, at least on this data set. On the other hand, PRISM had a high specificity rate having only 12 false positives. In other words, PRISM only predicted 38 adults without ASD traits that potentially supposed to be on the spectrum, and 12 individuals with ASD that are supposed to be classified as “No ASD.” The “No ASD” class has much higher data representation in the training data set than the ASD class, which means that the PRISM algorithm is sensitive to the number of data items linked to class labels; for RML, there were 15.

Figure 4(b) depicts the predictive accuracy of the considered ML algorithms including RML derived from the child, adolescent, and adult data sets. The figure clearly shows that the RML algorithm generated classifiers with higher accuracy than the considered algorithms on the adult and child data sets. The AdaBoost algorithm slightly outperformed RML on the adolescent data set, yet RML has derived a competitive classifier on the same data set.

The RIDOR algorithm has the largest number of false positives, wrongly predicting 33 instances having ASD who are supposed to be without ASD traits. Overall, there were higher classification rates for class “No ASD” than “ASD” with most of the considered algorithms. A probable reason for that fluctuation is that more instances representing class “No ASD” are present in the training data set. When the learning algorithm starts the training process, more rules are then derived for class “No ASD” in the classifier and therefore test instances that are supposed to be “No ASD” will have less misclassifications.

Since the adult autism data set is imbalanced with respect to class variable, researchers here included a metric called the harmonic mean (F1) that considers both recall (sensitivity) and precision (equation (6)). The F1 rates produced by the classifiers and shown in Figure 5 are high for the Covering (RML) and Boosting algorithms. This indicates that RML and AdaBoost perform well in data sets with imbalanced class labels and higher than the decision trees, Bagging, and rule induction algorithms represented by Nnge, RIDOR, RIPPER, CART, C4.5, and Bagging. For instance, on the adult data set, RML outperformed RIPPER, RIDOR, Bagging, CART, PRISM, and C4.5 with respect to F1 rate by 1.7, 3.1, 3.6, 2.2, 2.6, 1.8, and 1.6 percent, respectively.

The results produced by the ML algorithms with respect to error rate, sensitivity, specificity, and F1 reveal a promising direction for autism screening. The results also pinpointed that Covering algorithms, such as RML, work well in ASD detection for at least the adults. Furthermore, the



**Figure 6.** Number of rules derived by the considered ML algorithms on the adult data set.

performance of ML may be impacted when the number of instances for a class label is low, that is, class ASD. However, when a class is representative, such as “No ASD,” then the performance improves.

Overall, most algorithms generated acceptable sensitivity, specificity, and F1 results with more superiority to the Covering and Boosting classification approaches. These algorithms are more tolerant toward data with noise, that is, imbalanced data sets. A possible direction to boost the performance is to have more data for the low-frequency class.

The researchers investigated the classifier content generated by the Covering, decision tree, Bagging, and rule induction algorithms to seek important knowledge that could help in detecting ASD. Figure 6 shows the number of rules generated by the considered algorithms on the adult data set. The figures clearly show that the PRISM and Nnge algorithms generate the highest number of rules. The reason for extracting too many rules by PRISM is the fact that this algorithm has no rule pruning strategy, so it keeps building up rules, whereas Nnge is an algorithm that adopts the nearest neighbor search using non-nested generalized exemplars.

The number of rules results pinpointed that decision tree-like algorithms, such as CART and C4.5, derive classifiers larger in size than the rule induction and Covering approaches. The rule induction approach, represented by RIPPER and RIDOR, generates slightly smaller classifiers than the Covering approach. This can be attributed to the rigorous pruning procedures adopted by RIPPER and RIDOR in evaluating the rules.

Table 4 contains the common rules related to ASD detection that have been extracted by the Covering and rule induction approaches, respectively (RML, RIPPER). It seems that items in the AQ-Adult-10, screening methods have high influence on the class labels, particularly items 5, 9, 8, and 4. In addition, items 7 and 2 appeared in multiple rules in RIPPER and RML classifiers. It seems that the items that have frequently appeared within the rules cover certain autistic behaviors within the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5). For instance, item 8 covers repetitive behavior, item 4 is aligned with communication, and finally items 5 and 9 are aligned with social behavior (Table 5).

Finally, Table 6 depicts the runtime in milliseconds (ms) for the considered ML algorithms in processing the three autism data sets. It is obvious from the table that RML is more efficient than the remaining ML algorithms in building the models and for all data sets considered. Overall, ML techniques showed good efficiency in deriving the screening models from the child, adolescent, and adult data sets, respectively.

**Table 4.** Common rules derived by RML and RIPPER algorithms on the autism data sets.

RML rules (Freq, R_S)	RIPPER rules
1. (238, 1.00) Label=NO when A5_Score=0, A3_Score=0 2. (054, 1.00) Label=NO when A5_Score=0, A8_Score=0 3. (037, 1.00) Label=NO when A1_Score=0, A10_Score=0 4. (019, 1.00) Label=NO when A4_Score=0, A5_Score=0 5. (023, 1.00) Label=YES when A6_Score=1, A7_Score=1, family_with_autism=yes 6. 07 - (056, 1.00) Label=YES when A6_Score=1, A7_Score=1, A9_ Score=1, A1_Score=1 7. (039, 1.00) Label=NO when A4_Score=0, A2_Score=0, A9_Score=0 8. (030, 1.00) Label=YES when A6_Score=1, A3_Score=1, A1_ Score=1, A2_Score=1 9. (026, 1.00) Label=NO when A8_Score=0, A9_Score=0, born_ with_jaundice=no 10. 14 - (023, 1.00) Label=YES when A9_Score=1, A2_Score=1, A8_Score=1 11. (018, 0.86) Label=YES when A7_Score=1, A4_Score=1, A5_Score	Adult data set 1. If (A9_Score=1) and (A5_Score=1) and (A6_Score=1) and (A10_Score=1) => Class/ASD=YES (102.0/3.0) 2. If (A9_Score=1) and (A3_Score=1) and (A1_Score=1) and (A5_Score=1) => Class/ASD=YES (40.0/4.0) 3. If (A4_Score=1) and (A6_Score=1) and (A7_Score=1) and (A8_Score=1) => Class/ASD=YES (18.0/1.0) 4. If (A5_Score=1) and (A2_Score=1) and (A10_Score=1) and (A8_Score=1) and (A3_Score=1) => Class/ASD=YES (16.0/2.0) 5. If (A9_Score=1) and (A4_Score=1) and (A1_Score=1) and (A8_Score=1) and (A2_Score=1) => Class/ ASD=YES (8.0/0.0) 6. If (A7_Score=1) and (A5_Score=1) and (A4_Score=1) and (A8_Score=1) and (A10_Score=1) => Class/ ASD=YES (11.0/2.0) Adolescent data set 1. (A4_Score=0) and (A10_Score=0) => Class/ASD=NO 2. (A5_Score=0) and (A7_Score=0) => Class/ASD=NO 3. (A3_Score=0) and (A2_Score=0) => Class/ASD=NO Child data set 1. (A4_Score=1) and (A7_Score=1) and (A10_ Score=1) => Class/ASD=YES (79.0/1.0) 2. (A8_Score=1) and (A1_Score=1) and (A10_ Score=1) => Class/ASD=YES (35.0/6.0) 3. (A9_Score=1) and (A5_Score=1) and (A2_ Score=1) => Class/ASD=YES (17.0/1.0) 4. (A9_Score=1) and (A5_Score=1) and (A8_Score=1) and (A3_Score=1) => Class/ASD=YES (9.0/2.0) 5. (A4_Score=1) and (A5_Score=1) and (A1_Score=1) and (A3_Score=1) and (A2_Score=1) => Class/ ASD=YES (5.0/0.0) 6. (A9_Score=1) and (A4_Score=1) and (A10_Score=1) and (A3_Score=1) => Class/ASD=YES (4.0/0.0) Class/ ASD=NO (143.0/2.0)

**Table 5.** Common features mapping with AQ-Adult-10 screening method.

Item	Description
5	I find it easy to “read between the lines” when someone is talking to me
9	I find it easy to work out what someone is thinking or feeling just by looking at their face
8	I like to collect information about categories of things (e.g. types of car, types of bird, types of train, and types of plant)
4	If there is an interruption, I can switch back to what I was doing very quickly

**Table 6.** Time to build the models in milliseconds (ms).

	RIPPER	RIDOR	Nnge	Bagging	AdaBoost	CART	RML	C4.5	PRISM
Child	2	2	3	2	3	4	0	0	1
Adolescent	5	3	5	4	5	5	0	1	2
Adult	7	4	8	4	6	7	1	3	3

**Conclusion**

Autism screening is a fundamental step toward understanding autistic traits and for speeding up referrals to further evaluation in a clinical setting. However, existing screening tools such as AQ, Q-CHAT, and many others rely on simple calculation, using scoring functions that tally the scores of answers given by individuals. During the screening process, these scoring functions have been developed based on handcrafted rules and thus can be criticized for being subjective. Therefore, one of the crucial issues in ASD screening research is improving the screening process so that individuals and their families can have a faster and more accurate service. This can be accomplished by utilizing automated methods based on ML that build accurate classification systems from historical cases and controls. This article proposes a new ML method called RML that not only boosts the sensitivity, specificity, and predictive accuracy of the ASD screening process but also offers automatic classification besides rich rule sets for clinicians, caregivers, patients and their families, and teachers.

The proposed method generated non-redundant rules in a straightforward manner utilizing Covering learning. Empirical evaluation on different autism data sets using rule induction, Bagging, Boosting, and decision trees algorithms reported the superiority of the RML model with respect to different evaluation metrics including specificity, sensitivity, harmonic mean, and error rate. The results also showed that the RML-derived classifiers contain useful rules for understanding the reasons behind the ASD classification. Finally, the classifier’s content revealed some influential items in the autism screening that are aligned with social and communication behaviors yet not fully fulfilling the DSM-5 criteria for ASD diagnosis.

In conclusion, this article clearly revealed that ML approaches, especially Covering, showed promising results in detecting ASD cases especially for adults. In future, the researchers intend to design and implement a new autism screening tool based on rule sets derived by our model for toddlers, children, and adolescents.

One of the limitations of this article is not including instances related to toddlers as these are rare and difficult to obtain. In addition, RML possibly needs a method to deal with data sets that are imbalanced with respect to class labels, to further improve its predictive performance. Soon, we are going to build a new screening mobile application that will embed the rule-based classifiers to help diagnosticians access a rich knowledge base for improving screening and diagnostic decisions related to autism.

**Authors Note**

Fadi Thabtah is now affiliated with Digital Technologies, Manukau Institute of Technology.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.



## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Fadi Thabtah  <https://orcid.org/0000-0002-2664-4694>

## References

1. Towle P and Patrick P. *Autism spectrum disorder screening instruments for very young children: a systematic review*. New York: Hindawi Publishing Corporation, 2016.
2. Centers for Disease Control and Prevention. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR* 2014; 63(2): 1–21.
3. Lord C, Rutter M and Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994; 24: 659–685.
4. Lord C, Risi S and Lambrecht L. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 2000; 30: 205–223.
5. Fitzgerald M. The clinical gestalts of autism: over 40 years of clinical experience with autism. In: Fitzgerald M and Yip J (eds) *Autism: paradigms, recent research and clinical applications*. Rijeka: InTech, 2017, pp. 1–13.
6. Crane L, Chester J, Goddard L, et al. Experiences of autism diagnosis: a survey of over 1000 parents in the United Kingdom. *Autism* 2016; 20(2), 153–162.
7. Stone W, Coonrod E and Ousley O. Brief report: screening tool for autism in two-year-olds. *J Autism Dev Disord* 2000; 30: 607–612.
8. Schopler E and Bourgondien M. *(CARST<sup>TM</sup>-2) Childhood Autism Rating Scale<sup>TM</sup>, Second Edition*. Torrance, CA: WPS, 2010.
9. Schopler E, Reichler R and DeVellis R. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *J Autism Dev Disord* 1980; 10: 91–103.
10. Baron-Cohen S, Wheelwright S, Skinner R, et al. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* 2001; 31: 5–17.
11. Allison C, Auyeung B and Baron-Cohen S. Toward brief “red flags” for autism screening: the Short Autism Spectrum Quotient and the Short Quantitative Checklist for Autism in toddlers in 1,000 cases and 3,000 controls. *J Am Acad Child Psy* 2012; 51(2): 202–212.e7.
12. Wall DP, Dally R, Luyster R, et al. Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS ONE* 2012; 7: e43855.
13. Bone D, Goodwin MS, Black MP, et al. Applying machine learning to facilitate autism diagnostics: pitfalls and promises. *J Autism Dev Disord* 2014; 45(5): 1–16.
14. Duda M, Ma R, Haber N, et al. Use of machine learning for behavioral distinction of autism and ADHD. *Transl Psychiat* 2016; 9(6): e732.
15. Bone D, Bishop S, Black MP, et al. Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *J Child Psychol Psyc* 2016; 57: 927–937.
16. Thabtah F. Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment. In: *Proceedings of the 1st international conference on medical and health informatics 2017, Taichung City, Taiwan, 20–22 May 2017*, pp. 1–6. New York: ACM.
17. Pratap A, Kanimozhiselvi CS, Vijayakumar R, et al. Predictive assessment of autism using unsupervised machine learning models. *Int J Adv Intell Paradig* 2014; 6(2): 113–121.

18. Pancers K, Derkacz A and Gomula J. Consistency-based pre-processing for classification of data coming from evaluation sheets of subjects with ASDS. In: *Federated conference on computer science and information systems*, Lodz, 13–16 September 2015, pp. 63–67. Warsaw: Polskie Towarzystwo Informatyczne.
19. Kosmicki JA, Sochat V, Duda M, et al. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Transl Psychiat* 2015; 5: e514.
20. Levy S, Duda M, Haber N, et al. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Mol Autism* 2017; 8: 65.
21. Al-Diabat M. Fuzzy data mining for autism classification of children. *Int J Adv Comput Sci Appl* 2018; 9(7): 11–17.
22. Mohammad R, Thabtah F and McCluskey L. Intelligent rule-based phishing websites classification. *IET Inform Secur* 2014; 8(3): 153–160.
23. Thabtah F. Review on associative classification mining. *Knowl Eng Rev* 2007; 22(1): 37–65.
24. Chang CC and Lin CJ. LIBSVM: a library for support vector machines. *ACM T Intel Syst Tec* 2011; 2: 27.
25. Quinlan J. Induction of decision trees. *Mach Learn* 1986; 1(1): 81–106.
26. Cohen W. Fast effective rule induction. In: *Proceedings of the twelfth international conference on machine learning*, Tahoe City, CA, 9–12 July 1995, pp. 115–123. San Francisco, CA: Morgan Kaufmann.
27. Freund Y and Schapire RE. A decision-theoretic generalization of online learning and an application to boosting. *J Comput Syst Sci* 1997; 55: 119–139.
28. Breiman L. Bagging predictors. *Mach Learn* 1996; 24: 123–140.
29. Mohammad R, Thabtah F and McCluskey L. An improved self-structuring neural network. In: *Pacific-Asia conference on knowledge discovery and data mining*, Auckland, New Zealand, 19–22 April 2016, pp. 35–47. Cham: Springer.
30. Qabajeh I, Thabtah F and Chiclana F. A dynamic rule-induction method for classification in data mining. *J Manage Anal* 2015; 2(3): 233–253.
31. Thabtah F. ASDTests: a mobile app for ASD screening, [www.asdtests.com](http://www.asdtests.com) (2017, accessed 30 November 2017).
32. Lichman M. *UCI machine learning repository*. Irvine, CA: School of Information and Computer Science, University of California, 2013, <http://archive.ics.uci.edu/ml>
33. Lopez Marciano JL. *Classification of ADHD and non-ADHD using AR models and machine learning algorithms*. Doctoral Dissertation, Virginia Tech, Blacksburg, VA, 2016.
34. Thabtah F. Machine learning in autistic spectrum disorder behavioral research: a review. *Inform Health Soc Ca*. Epub ahead of print 13 February 2018. DOI: 10.1080/17538157.2017.1399132.
35. Thabtah F, Kamalov F and Rajab K. A new computational intelligence approach to detect autistic features for autism screening. *Int J Med Inform* 2018; 117: 112–124.
36. Ruzich E, Allison C, Smith P, et al. Measuring autistic traits in the general population: a systematic review of the autism-spectrum quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Mol Autism* 2015; 6: 2.
37. Wolfers T, Buitelaar JK, Beckmann CF, et al. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci Biobehav R* 2015; 57: 328–349.
38. Wall DP, Kosmicki J, Deluca TF, et al. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl Psychiat* 2012; 2: e100.
39. Geschwind DH, Sowinski J, Lord C, et al. The Autism Genetic Resource Exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet* 2001; 69: 463–466.
40. Hall M, Frank E, Holmes G, et al. The WEKA data mining software: an update. *SIGKDD Explorations* 2009; 11(1): 10–18.
41. Fischbach GD and Lord C. The Simons simplex collection: a resource for identification of autism genetic risk factors. *Neuron* 2010; 68: 192–195.
42. Rutter M, Bailey A and Lord C. *The Social Communication Questionnaire Manual*. Torrance, CA: WPS, 2003.

43. Chu KC, Huang HJ and Huang YS. Machine learning approach for distinction of ADHD and OSA. In: *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, San Francisco, CA, 18–21 August 2016, pp. 1044–1049. New York: IEEE.
44. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Publishing, 2000.
45. Achenbach T. *Manual for the youth self-report and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont, 1991.
46. Maenner MJ, Yeargin-Allsopp M, Van Naarden Braun K, et al. Development of a machine learning algorithm for the surveillance of autism spectrum disorder. *PLoS ONE* 2016; 11(12): e0168224.
47. Breiman L. Random forests. *Mach Learn* 2001; 45(1): 5–32.
48. Thabtah F, Qabajeh I and Chiclana F. Constrained dynamic rule induction learning. *Expert Syst Appl* 2016; 63: 74–85.
49. Agrawal R, Imieliński T and Swami A. Mining association rules between sets of items in large databases. In: *Proceedings of the conference on management of data*, Washington, DC, 25–28 May 1993, pp. 207–216. New York: ACM.
50. Thabtah F. An accessible and efficient autism screening method for behavioural data and predictive analyses. *Health informatics journal*, Epub ahead of print 19 September 2018. DOI: 10.1177/1460458218796636.
51. Gaines BR and Compton P. Induction of ripple-down rules applied to modeling large databases. *J Intell Inf Syst* 1995; 5(3): 211–228.
52. Martin B. *Instance-based learning: nearest neighbor with generalization*. Hamilton, New Zealand: University of Waikato, 1995.
53. Breiman L, Friedman JH, Olshen RA, et al. *Classification and regression trees*. Belmont, CA: Wadsworth International Group, 1984.
54. Quinlan J. *C4.5: programs for machine learning*. San Mateo, CA: Morgan Kaufmann, 1993.
55. Cendrowska J. PRISM: an algorithm for inducing modular rules. *Int J Man Mach Stud* 1987; 27(4): 349–370.
56. Witten IH and Frank E. *Data mining: practical machine learning tools and techniques*. San Francisco, CA: Morgan Kaufmann, 2005.
57. Abdelhamid N and Thabtah F. Associative classification approaches: review and comparison. *J Inform Knowl Manag* 2014; 13(3): 1450027.