Pertanika J. Sci. & Technol. 34 (2): 695 - 720 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Association Rule Mining to Detect Significant Factors of Autism Spectrum Disorder at Early Stage

Tania Akter[1,2], Farzana Tasnim[2], Shahriar Hassan[2], Mohammad Hanif Ali[1], and Mohammad Shorif Uddin[1]*

[1]Department of Computer Science and Engineering, Jahangirnagar University, 1342 Savar, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Gono Bishwabidyalay, 1344 Savar, Dhaka, Bangladesh

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex, long-lasting neurodevelopmental condition characterised by communication difficulties. However, early identification of ASD may aid in the design of appropriate treatments to improve communicative development. This study employs association rule mining algorithms, such as apriori, predictive apriori, and Tertius, to identify age-specific behavioural markers across four developmental stages: toddlers, children, adolescents, and adults. These algorithms identified several key behavioural indicators associated with ASD, including pretending games (A5), activity switching (A4), difficulty in making friends (A10), difficulty in conversations (A5), detection of listener boredom (A6) and easy reading of emotions (A9). The apriori algorithm achieved the highest confidence of 100.00% for the adolescent dataset, whereas the predictive apriori algorithm demonstrated the highest accuracy of 99.50% for the toddler dataset, 99.40% for the child dataset and 99.39% accuracy for adult dataset. In contrast, the Tertius algorithm showed the highest confirmation rate of 66.8% for the child dataset, although it required the most time for manipulation across all datasets. Our findings demonstrate that rule mining effectively uncovers clinically relevant patterns, with behavioural questions proving more significant than demographic factors, such as gender or birth complications. Furthermore, no connection was observed between ASD and a history of jaundice at birth. Performance comparisons among the three algorithms are also presented. The rules generated through our investigation will help physicians in the early detection of ASD, thus paving the way for a timely and targeted intervention.

Keywords: AQ-10 screening dataset, association rule mining, autism spectrum disorder (ASD), Q-Chat-10 toddler

ARTICLE INFO

Article history:
Received: 13 September 2024
Accepted: 15 December 2025
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.05

E-mail addresses:
shilpy339cse@gmail.com (Tania Akter)
rekha.tasnim@gmail.com (Farzana Tasnim)
shahriar.hassan303@gmail.com (Shahriar Hassan)
hanif_ju03@juniv.edu (Mohammad Hanif Ali)
shorifuddin@juniv.edu (Mohammad Shorif Uddin)
* Corresponding author

e-ISSN: 2231-8526

INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised by a deficiency of social behaviour and interaction (Mazumdar et al., 2021; Rubio-Martín et al., 2024; Uddin et al., 2023). Currently, the number of people with ASD is rapidly increasing, with almost 1 in every 68 children diagnosed with it. Therefore, approximately 1.5% of the entire population suffers from ASD (Towle & Patrick, 2016; Wingate et al., 2014), which has become a concern. Although Autism Spectrum Disorder (ASD) cannot be completely cured, research has shown that early intervention can substantially improve developmental progress and long-term outcomes. However, obtaining a timely diagnosis remains a significant challenge due to the diverse range of symptoms and the dependence on subjective behavioural assessments. Structured tools for monitoring progress, such as screening questionnaires, play a vital role in identifying early risk indicators. However, the complexity of multidimensional data analysis often impedes the generation of actionable insights.

Association Rule Mining (ARM) techniques (Agrawal et al., 1993; Kaoura et al., 2024) are used to extract significant correlations, frequent patterns, associations, or causal structures between all items in datasets (Zhao & Bhowmick, 2003). By analysing behavioural and demographic variables in the ASD screening data, the ARM can recognise combinations of risk factors that are strongly related to ASD diagnosis, allowing for more precise and objective early detection. Despite its potential, ARM remains underexplored in ASD research compared to conventional machine learning approaches.

The application of association rule mining (ARM) methods to the diagnosis of autism spectrum disorder (ASD) has recently received growing attention (Singh & Mantri, 2024). Several studies have investigated the use of ARM to detect patterns related to autism characteristics and probable risk factors. For example, Thabtah and Peebles (2020) proposed a new rule-based machine learning method that detected autism traits and extracted significant rules. In addition, Al-Diabat (2018) evaluated a fuzzy rule-based data mining method for forecasting autistic symptoms in children that can be automatically extracted from previous instances to form a screening classification system. Wiratsin and Narupiyakul (2021) proposed a novel feature selection method to identify important relationships between ASD symptoms in three different age groups. Engle and Rada (2011, 2017) used association rule mining techniques to generate rules. Engle (2013) also proposed a novel approach to combine domain knowledge with association rule mining methods. In addition to association rule mining, many researchers have also worked using various machine learning algorithms such as Random Forest, Support Vector Machine, Gradient Boost, Neural Networks, etc. to the mentioned datasets. In their work, they showed high accuracy to predict autism, ranging from approximately 85% - 98% in these datasets (Akter et al., 2021; Uddin et al., 2023). But they could not show which behaviours were especially responsible for detecting autism. But in the case of association rules, these algorithms identify important specific behavioural markers and explained relationship among them. For small datasets, association rule mining is better than machine learning in terms of interpretability and feature discovery.

In previous work, researchers have used ARM in the child, adolescent, and adult data sets or used a single cohort to find key factors related to ASD, but no one has used these techniques in the Q-Chat-10 toddler data set yet. In our study, we conducted a comparative and multicohort study of four developmental age stages, such as toddler, child, adolescent, and adult, in which the same parameters of association rule mining were utilised. This analysis provides a broad perspective on behavioural development and the importance of these factors.

While existing research has made contributions, there is a scarcity of studies that specifically utilise association rule mining for ASD diagnosis. This study seeks to fill this research gap by applying association rule mining techniques to analyse various ASD datasets, including the Q-Chat-10 toddler, AQ-10 child, AQ-10 adolescent, and AQ-10 adult screening datasets. Through the implementation of ARM methodologies, we derived rules that differentiate ASD from non-ASD conditions, aiming to identify the most effective strategies for early detection. This study is anticipated to offer crucial insights for medical professionals and caregivers, enabling simpler and more precise identification of ASD.

The vital contributions of our work are mentioned below:

- Various risk factors of ASD and non-ASD individuals are identified using association rule mining algorithms.
- Associations among different risk factors that indicate ASD have been mined using different association rule mining algorithms.
- The performance of different association rule mining algorithms is compared.

The rest of the paper goes through the following organised. Section II presents the details of the datasets used and the methodology of this work. In Section III, the extracted rules are presented, and the performance analysis is provided. A thorough discussion of the findings is provided in Section IV. Lastly, Section V concludes the work.

MATERIALS AND METHODS

In this study, we employed toddlers, children, adolescents, and adults to analyse ASD using the association rule mining technique. Next, we pre-processed these datasets and applied apriori, predictive apriori, and tertius algorithms to uncover relationships and patterns within the data. The rules identified through this analysis were further explored to determine the most significant rules that could assist in early diagnosis and interventions. To detect risk factors for ASD, the basic pipeline of this study is shown in Figure 1.

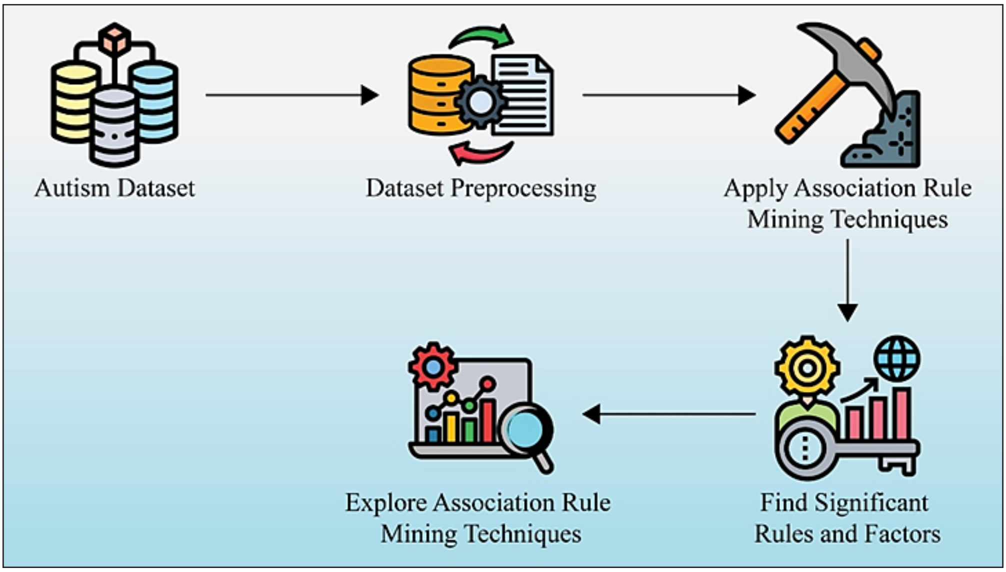

Figure 1. Association rule mining framework for identifying key factors in autism spectrum disorder

Dataset Description

To detect significant factors of autism, we gathered Q-Chat 10 toddler (Delja, n.d.), AQ-10 child, AQ-10 adolescent, and AQ-10 adult datasets from Kaggle and UCI data repositories, respectively, which are created based on AQ-10 screening tool questionnaires (UCI Machine Learning Repository, n.d.-a, -b, -c). These data have been collected using a mobile application called ASDTests. All datasets contained 20 features, including class labels. This label contains two values, one is ”0”, which indicates the individuals with no ASD traits, and ”1”, which indicates that they contain ASD traits. The toddlers screening dataset contains 1054 cases in the age group 18-36 months where 30.26% female and 69.73% male, the child screening dataset contains 292 cases in the age group 4-11 years where 28.77% female and 71.23% male, the adolescent screening dataset contains 104 cases with the age group 12-16 years where 51.92% female and 48.08% male and adult screening dataset contains 704 cases with the age group above 18 years where 47.87% female and 52.13% male. Thus, the toddler and child datasets are not balanced, but the adolescent and adult datasets are balanced (Figure 2).

Figure 3 illustrates the distribution of ASD cases based on the presence of jaundice at birth across all four datasets. The number of cases with jaundice was lower than that without jaundice, suggesting that there was no strong association between jaundice at birth and ASD occurrence.

Figure 4 shows the distribution of ASD cases based on family ASD across all four datasets. In all datasets, the number of cases without a family history of ASD was greater than that of cases with a family history. This suggests that there is no strong association between family ASD and the occurrence of ASD. The description of the dataset is shown in Table 1.

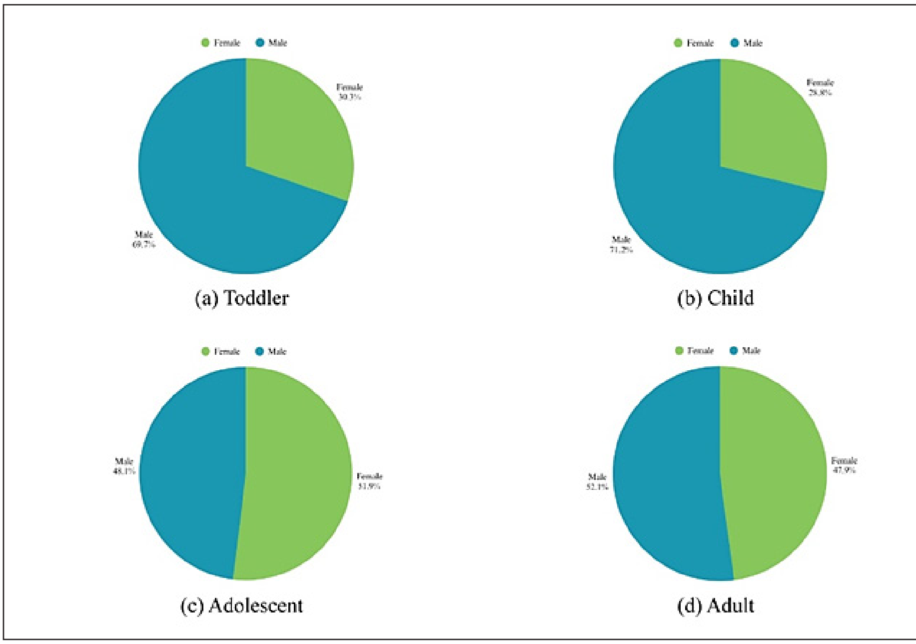

Figure 2. Gender based distribution of ASD cases

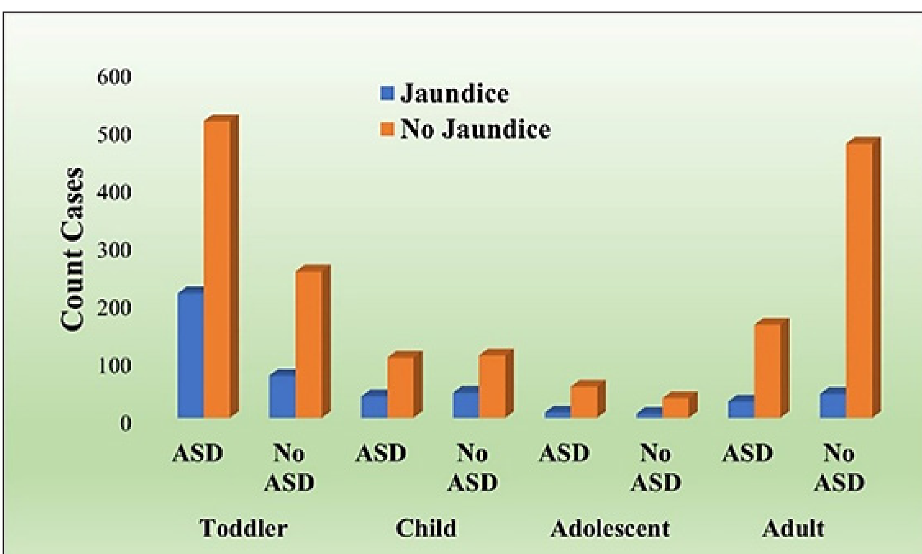

Figure 3. Association between jaundice at birth and autism spectrum disorder

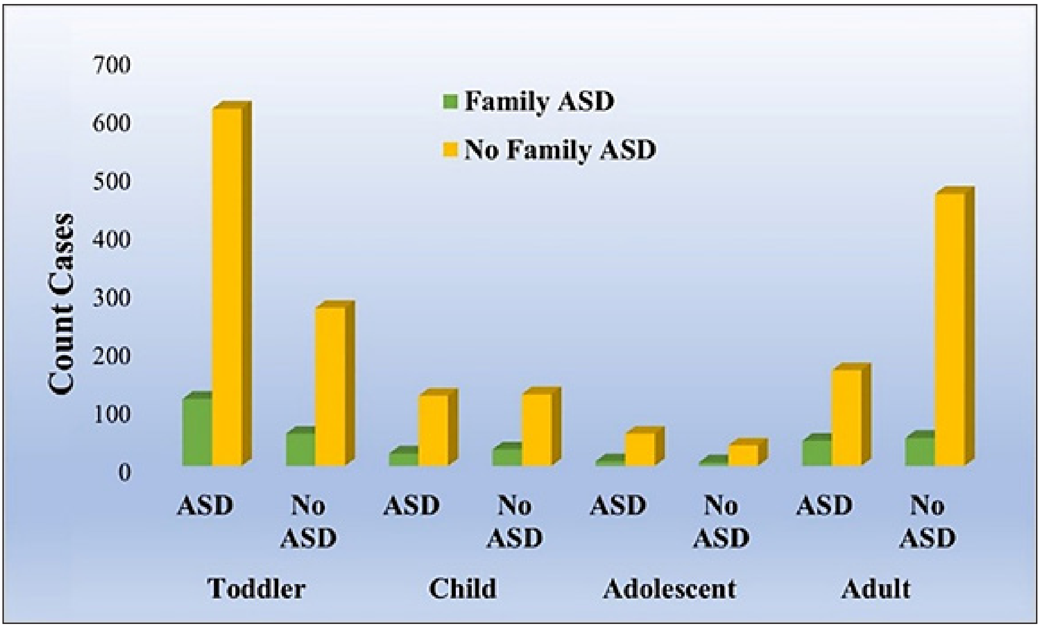

Figure 4. Impact of family ASD on offspring's autism spectrum disorder

Table 1
Features description

Feature	Type	Description
Age	Number	toddlers (month), children, adolescents, and adults(year)
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean	Whether the case was born with jaundice
Family member with ASD	Boolean	Whether any immediate family member has an ASD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician, etc.
Country of residence	String	List of countries in text format
Used the screening app before	Boolean	Whether the user has used a screening app
Screening Method Type	Integer	The type of screening methods chosen is based on the age category
A1	Binary	The response to the question depends on the screening method employed.
A2	Binary	The response to the question depends on the screening method employed.

Table 1 (*continued*)

Feature	Type	Description
A3	Binary	The response to the question depends on the screening method employed.
A4	Binary	The response to the question depends on the screening method employed.
A5	Binary	The response to the question depends on the screening method employed.
A6	Binary	The response to the question depends on the screening method employed.
A7	Binary	The response to the question depends on the screening method employed.
A8	Binary	The response to the question depends on the screening method employed.
A9	Binary	The response to the question depends on the screening method employed.
A10	Binary	The response to the question depends on the screening method employed.
Scoring Result	Integer	The response to the question depends on the screening method employed.
ASD	Boolean	toddlers, children, adolescent or adults diagnosed with ASD

Data Preprocessing

After collecting the primary datasets, we found that they contained several missing values. If the data are numeric, the missing value is generally replaced by the mean/median. In our dataset, the nominal attributes contained missing values. Therefore, we replace it with a constant (Filter: ReplaceMissing WithUserConstant). This filter works as follows: if any nominal attribute contains two classes and class one contains 90 per cent of the data, then this filter replaces the missing value with the constant value of class one. We also converted the numerical data into nominal values. The correlation method was then used to eliminate highly correlated features from these datasets (Akter et al., 2021). Therefore, QChat10 Score is eliminated from the toddlers' dataset, age desc, and results are removed from the child and adolescent datasets, and the result attribute is reduced from the adult dataset.

Association Rule Mining

In this study, association rule mining was used to analyse the data from autism spectrum disorder. Every association rule mining method has a basic structure to fulfil the minimum support and confidence (Ishibuchi et al., 2007). To extract significant features of autism more accurately, different association rule mining algorithms, such as apriori, predictive apriori, and tertius, were employed in all autism datasets in this study. These algorithms have been extensively used to analyse the significant factors of different datasets (Nahar et al., 2013; Niranjan et al., 2016).

Additionally, these algorithms have their complementary analytical properties. The apriori algorithm has high efficiency to identify confident frequent patterns, the predictive apriori has the capability to use probabilistic optimisation for improved accuracy, and the tertius has capability to extract confirmatory and relational rules. By using them together, we can get a balanced assessment through confidence, accuracy, and confirmation metrics.

Apriori Algorithm

The apriori algorithm (Cai, 2020; Maitra et al., 2021) is an association rule mining technique that employs frequent itemsets to create association rules. This algorithm calculates the level of relationship among objects using the association rule and works well on transactional databases. To estimate the associations of items proficiently, a breadth-first search and a hash tree are applied. In this algorithm, the support of the itemset is first calculated, and the minimum support and confidence are fixed. Next, all item sets whose support values are greater than the minimum support value is selected. Finally, rules for subsets with higher confidence values than the minimum confidence value are created. This algorithm is easy to understand, and a complex formula does not need to be derived. However, it works gradually, and its time and space complexities are high. In the apriori algorithm, the rules are ranked by the confidence, like the accuracy value of the rules (Mutter et al., 2004; Wang & Guo, 2010).

Predictive Apriori Algorithm

The predictive apriori algorithm, first introduced by Scheffer (2005), builds upon the traditional apriori algorithm. Unlike apriori, which generates association rules based solely on confidence (Aher & Lobo, 2012), predictive apriori considers both support and confidence to rank rules more effectively. Within a Bayesian framework, it estimates the expected accuracy of a rule using its support and confidence, thereby enhancing its ability to generalise the results. When applied to a database D with a set of records r and a static process s, the predictive accuracy of the rule: I >= J is the conditional probability of J being a subset of r, where I is a subset of r (Nahar et al., 2013).

Tertius Algorithm

Tertius is an algorithm that finds the rule according to the measurement of confirmation and a top-down rule-finding system using first-order logic representation to deal with structured, relational information (Flach & Lachiche, 2001; Hareendran & Chandra, 2017). Several confirmation measures have been examined, with the simplest being weighted relative accuracy. A confirmation measure reveals the degree to which a rule is unexpected, as well as the percentage of expected counter-instances. The algorithm computes two values: the expected probability and the observed probability (Nahar et al., 2013).

Evaluation Metrics for Association Rule Mining

In this work, we justified the corresponding rules by considering some statistical evaluation metrics, such as confidence (Kane, 2018; Yan et al., 2009), accuracy, confirmation, and support (Hamid et al., 2024) of ASD and non-ASD. This analysis reflects the usefulness and certainty of a rule using these metrics.

When we consider a rule ($X \Rightarrow Y$),

$$Confidence\ (X \Rightarrow Y) = P(Y|X) = \frac{(Support\ \ (X \cap Y))}{(Support\ \ (X))} \quad [1]$$

$$Support\ (X \Rightarrow Y) = P(X \cap Y) \quad [2]$$

where $P(X)$ is defined as the probability of randomly selected transactions of the dataset including all items in X and $X \cap Y$ is indicated as all transactions which contained both X and Y (Hasanpour et al., 2019). On the other hand, support and confidence are used as a single measure called accuracy, which is denoted as follows:

$$Accuracy = \{Support, Confidence\} \quad [3]$$

Then, confirmation is used to measure the novelty of the associated rules. Calculating results requires expected and observed frequency values.

$$Confirmation = \frac{expected-observed}{\sqrt{expected-expected}} \quad [4]$$

To make sure the reproducibility of our results, we applied a consistent minimum support and confidence threshold to all age groups. In the case of apriori algorithm, the minimum support is set at 0.05, and the confidence threshold is set at 0.90. In the case of apriori predictive algorithm, the minimum support is set at 0.04, and the confidence threshold is set at 0.85. For the tertius algorithm, the confirmation value is set at 0.75. Such threshold values were chosen to balance the quality of the rules and the number of rules in all the datasets.

RESULTS

We used the three most prevalent association rule mining algorithms, named apriori, predictive apriori, and tertius, to conduct this study using Weka 3.8 (Aksu & Dogan, 2019).

These algorithms generate several rules, but we only considered those containing class attributes on the right-hand side or left-hand side. Each rule contained a set of features linked to the ASD outcome, accompanied by confidence accuracy, and confirmation scores representing the rule's reliability. The most significant rules were reported based on confidence, accuracy, and confirmation. Tables 2 to 13 represent the results of the experiment. In all results tables, the symbol "∩" denotes logical AND (intersection), while "∪" denotes logical OR (union) between conditions in association rules. (A1- A9) are personal behavioural questions on the ASD screening data. The score of 1 represents the presence of the behaviour, whereas 0 represents the absence of the behaviour.

Results for Identifying ASD in Toddlers

Table 2 illustrates the association rules generated by the apriori algorithm utilised for a dataset of toddlers diagnosed with ASD. Each rule contained a set of features linked to the ASD outcome, accompanied by confidence scores representing the rule's reliability. For instance, Rule 1 (A5 Score=1 ∩ A6 Score=1 → ASD=Yes) has a confidence of 0.990, indicating that 99.00% of the time, when a toddler scores 1 on both A5 and A6, they tend to have ASD. All rules showed the confidence value above 0.920, reflecting robust relationships with ASD. The analysis reveals robust correlations among several attributes, including A1, A2, A5, A6, A7, and A9. This algorithm gave no such rules for healthy toddlers.

Table 3 presents association rules derived from the predictive apriori algorithm, revealing patterns related to ASD diagnosis in toddlers. These rules demonstrate high accuracy, ranging from 0.9940 to 0.9950, indicating strong predictive capability. For example, Rule 1 states that if A4, A5, and A9 scores are all 1, there is a 99.50% probability

Table 2
Association rules generated by the apriori algorithm for detecting autism spectrum disorder in toddlers

Rule No.	Association Rules	Confidence
1	A5_Score=1 ∩ A6_Score=1 → ASD=Yes	0.990
2	A5_Score=1 ∩ A7_Score=1 → ASD=Yes	0.980
3	A1_Score=1 ∩ A7_Score=1 → ASD=Yes	0.970
4	A1_Score=1 ∩ A6_Score=1 → ASD=Yes	0.970
5	A6_Score=1 ∩ A7_Score=1 → ASD=Yes	0.960
6	A9_Score=1 → ASD=Yes	0.960
7	A5_Score=1 → ASD=Yes	0.940
8	A5_Score=1 ∩ autism=0 → ASD=Yes	0.930
9	A6_Score=1 ∩ autism=0 → ASD=Yes	0.930
10	A2_Score=1 → ASD=Yes	0.930

Table 3
Association rules generated by the predictive apriori algorithm for detecting autism spectrum disorder in toddlers

Rule No.	Association Rules	Accuracy
1	A4_Score=1∩A5_Score=1∩A9_Score=1→ASD=Yes	0.9950
2	A5_Score=1∩A9_Score=1∩Sex= male →ASD=Yes	0.9950
3	A1_Score=1∩A5_Score=1∩A9_Score=1→ASD=Yes	0.9950
4	A3_Score=1∩A4_Score=1∩A7_Score=1→ASD=Yes	0.9950
5	A1_Score=1∩A6_Score=1∩A10_Score=1→ASD=Yes	0.9950
6	A3_Score=1∩A4_Score=1∩A6_Score=1→ASD=Yes	0.9950
7	A1_Score=1∩A4_Score=1∩A5_Score= 1 → ASD =Yes	0.9950
8	A1_Score=1∩A5_Score=1∩A6_Score=1∩A7_ Score=1→ASD=Yes	0.9949
9	A2_Score=1∩A4_Score=1∩A7_Score=1→ASD=Yes	0.9949
10	A5_Score=1∩A6_Score=1∩A8_Score=1→ASD=Yes	0.9949
11	A5_Score=0∩A6_Score=0∩A9_Score=0→ASD=No	0.9946
12	A3_Score=0∩A5_Score=0→ASD=No	0.9944
13	A4_Score=0∩A5_Score=0→ASD=No	0.9943
14	A3_Score=0∩A4_Score=0∩A6_Score=0→ASD=No	0.9943
15	A5_Score=0∩A7_Score=0∩A9_Score=0∩jundice=0→ASD=No	0.9940

of ASD being present. Rule 11, on the other hand, indicates that if A5, A6, and A9 scores are all 0, the probability of ASD being absent is 99.46%. These rules emphasise the critical attribute combinations that are strongly associated with ASD, providing valuable insights for early diagnosis and intervention.

Table 4 highlights the association rules obtained through the Tertius algorithm, which identifies additional patterns related to ASD in toddlers. The rules show that certain behavioural scores, like A5 and A9, in combination with factors such as age or ethnicity, are linked to ASD diagnosis. For instance, toddlers aged 16 or 26 months, or those from the Pacific ethnicity, who had A5 and A9 scores of 1, were more likely to be diagnosed with ASD, with confirmation scores ranging from 0.6043 to 0.6079. In contrast, toddlers with A2, A5, and A9 scores of 0 were less likely to have ASD. The highest confirmation score (0.6137) was associated with the rule A2_Score=0 ∩ A5_Score=0 ∩ A9_Score=0 → ASD=No, indicating a moderately reliable predictive association for the absence of ASD. Across all rules, confirmation scores ranged from 0.6042 to 0.6137, indicating moderate predictive strength. Comparing the three algorithms, it becomes evident that positive responses to pretending games (A5) and tracking vision (A6) are indicative of ASD presence, while negative responses to pretending games (A5) and simple gestures (A9) suggest the absence of ASD.

Table 4
Association rules generated by the tertius algorithm for detecting autism spectrum disorder in toddlers

Rule No.	Association Rules	Confirmation
1	A5_Score=1∪A9_Score=1∪Age_Mons=26→ ASD=Yes	0.6079
2	A5_Score=1∪A9_Score=1 ∪ Ethnicity = Pacifica → ASD=Yes	0.6057
3	A5_Score=1∪A9_Score=1∪Age_Mons=16→ ASD=Yes	0.6043
4	A2_Score=0∩A5_Score=0∩A9_Score=No→ ASD=No	0.6137
5	A5_Score=0∩A9_Score=0→ASD=No	0.6071
6	A2_Score=0∩A9_Score=0→ASD=No	0.6055

Results for Identifying ASD in Children

Table 5 presents the association rules generated by the apriori algorithm to identify the patterns associated with ASD in children. The results indicate that specific combinations of characteristics, such as positive responses to A3, A4, A5, and A10 (Rule 1), show the highest confidence level (95.00%) in predicting ASD. Rules 2–10 further highlight other feature pairings with confidence levels of 92.00%–94.00%. Notably, the algorithm did not generate any rules for the healthy children.

Table 6 shows the association rules derived from the predictive apriori algorithm, which demonstrates a high level of predictive accuracy, illustrating the algorithm's effectiveness in distinguishing between ASD and non-ASD cases. Rule 1, with positive responses for A1, A3, A4, A5, and A10, offers the highest accuracy of 99.44% for predicting ASD. Rule 2, which includes A3, A4, A9, and A10, follows closely with an accuracy of 99.43%. Additionally, several rules accurately predicted non-ASD cases by identifying the absence of certain attributes, such as A1, A4, and A9 in Rule 3 (A1_Score=0 ∩ A4_Score=0 ∩ A9_Score=0 → ASD=No), which achieved an accuracy of 99.34%.

According to Table 7, the tertius algorithm identified A4 and A10 scores as significant factors for predicting ASD in children, with a confirmation rate of 66.86%. To predict absence of ASD, the key indicators included A4_Score=0, A10_Score=0, or the child's country of residence (e.g., United Arab Emirates), with a confirmation rate of 67.56%. These factors were consistently present across the generated association rules, underscoring their importance in distinguishing between ASD and non-ASD patients. When comparing the three algorithms, it becomes clear that a positive response for quick activity swapping (A4) and difficulty in making friends (A10) can signal the presence of ASD, while negative responses for these same factors can indicate the absence of ASD.

Table 5
Association rules generated by the apriori algorithm for detecting autism spectrum disorder in children

Rule No.	Association Rules	Confidence
1	A3_Score=1∩A4_Score=1∩A5_Score=1∩A10_ Score=1→ASD=Yes	0.950
2	A1_Score=1∩A4_Score=1∩A5_Score=1→ASD =Yes	0.940
3	A1_Score=1∩A4_Score=1∩A10_Score=1→ASD =Yes	0.940
4	A3_Score=1∩A4_Score=1∩A5_Score=1∩A6_ Score=1→ASD=Yes	0.930
5	A4_Score=1∩A5_Score=1∩A6_Score=1∩A10_ Score=1→ASD=Yes	0.930
6	A4_Score=1∩A5_Score=1∩A10_Score=1→ASD =Yes	0.920
7	A3_Score=1∩A4_Score=1∩A6_Score=1∩A10_ Score=1→ASD=Yes	0.920
8	A5_Score=1∩A9_Score=1∩A10_Score=1→ ASD=Yes	0.920
9	A3_Score=1∩A4_Score=1∩A10_Score=1→ ASD=Yes	0.920
10	A1_Score=1∩A3_Score=1∩A5_Score=1∩A10_ Score=1→ASD=Yes	0.920

Table 6
Association rules generated by the predictive apriori algorithm for detecting autism spectrum disorder in children

Rule No.	Association Rules	Accuracy
1	A1_Score=1∩A3_Score=1∩A4_Score=1∩A5_ Score =1∩A10Score=1→ASD=Yes	0.9944
2	A3_Score=1∩A4_Score=1∩A9_Score=1∩A10_ Score=1→ASD=Yes	0.9943
3	A1_Score=0∩A4_Score=0∩A9_Score=0→ASD=No	0.9934
4	A1_Score=0∩A4_Score=0∩A8_Score=0→ASD=No	0.9933
5	A4_Score=0∩A10_Score=0→ASD=No	0.9932
6	A1_Score=0∩A10_Score=0→ASD=No	0.9924
7	A3_Score=0∩A6_Score=0==>Class/ASD=0→ ASD=No	0.9922
8	A5_Score=0∩A10_Score=0→ASD=No	0.9922
9	A1_Score=0∩A2_Score=0∩A8_Score=0→ASD=No	0.9915
10	A3_Score=0∩A7_Score=0→ASD=No	0.9915

Table 7
Association rules generated by the tertius algorithm for detecting autism spectrum disorder in children

Rule No.	Association Rules	Confirmation
1	A4_Score=1∩A10_Score=1→ASD=Yes	0.6686
2	A3_Score=1∩A4_Score=1∩A10_Score=Yes→ ASD=Yes	0.6651
3	A4_Score=0∪A10_Score=0 ∪ country_of_res= United_Arab_ Emirates →ASD=No	0.6756
4	A4_Score=0 ∪ A10_Score=0 ∪ country_of_res = Italy → ASD=No	0.6756
5	A4_Score=0 ∪ A10_Score=0 ∪ country_of_res = Bangladesh → ASD=No	0.6688
6	A4_Score=0 ∪ A10_Score=0→ASD=No	0.6686
7	A3_Score=0 ∪ A4_Score=0 ∪ A10_Score=0→ASD=No	0.6651
8	A4_Score=0 ∪ A10_Score=0 ∪ country_of_res = Jordan → ASD=No	0.6623
9	A4_Score=0 ∪ A10_Score=0 ∪ country_of_res = Russia → ASD=No	0.6620

Results for Identifying ASD in Adolescents

Table 8 illustrates that the apriori algorithm identified key characteristics strongly correlated with ASD in adolescents, particularly when A3, A5, A6, and A9 scores were all labelled as "Yes", resulting in a 100.00% confidence level. Other rules also showed significant correlations, with confidence levels ranging from 94.00% to 96.00%. Notably, the algorithm did not generate any rules for healthy adolescents, suggesting that these trends are specific to those with ASD and could serve as critical markers for early diagnosis and treatment.

Table 8
Association rules generated by the apriori algorithm for detecting autism spectrum disorder in adolescents

Rule No.	Association Rules	Confidence
1	A3_Score=1∩A5_Score=1∩A6_Score=1∩A9_ Score=1→ASD=Yes	1.000
2	A3_Score=1∩A5_Score=1∩A6_Score=1→ASD =Yes	0.960
3	A3_Score=1∩A5_Score=1∩A9_Score=1→ASD =Yes	0.960
4	A5_Score=1∩A6_Score=1∩A9_Score=1∩A10_ Score=1 → ASD=Yes	0.960
5	A3_Score=1∩A4_Score=1∩A5_Score=1→ASD =Yes	0.960
6	A4_Score=1∩A5_Score=1∩A8_Score=1→ASD =Yes	0.950
7	A5_Score=1∩A6_Score=1∩A10_Score=1→ASD=Yes	0.940

For the predictive apriori algorithm, as shown in Table 9, the primary factors for identifying ASD in adolescents were the A3, A5, A6, and A9 scores being "Yes", achieving an accuracy of 99.46%. In contrast, for healthy adolescents, the absence of A3 and A10 scores (both marked as "No") was a key indicator, with an accuracy of 99.30%. These findings highlight the clear distinctions between ASD and non-ASD adolescents, offering valuable insights into accurate classification and early intervention.

Table 10, featuring the tertius algorithm, shows that for adolescents with ASD, the significant factors (with a confirmation of 66.52%) were the A3, A5, and A6 scores being "Yes". For healthy adolescents (with a confirmation rate of 69.09%), key indicators included A6_Score=0, A10_Score=0, or the adolescent being 13 years old, with a confirmation rate of 69.09%. A comparison of the three algorithms revealed that a positive response for A3 (tracking conversation) and A5 (conversation difficulty) could indicate the presence of ASD, while a negative response to tracking conversation (A3) and difficulties with conversations (A5) could indicate that the adolescent was non-ASD.

Results for Identifying ASD in Adults

In Table 11, the apriori algorithm identified key factors indicative of healthy adults with a confidence level of 97.00%, particularly when A6_Score, A9_Score, and jaundice status were "No". These combinations strongly suggest the absence of ASD (ASD=No).

Table 9
Association rules generated by the predictive apriori algorithm for detecting autism spectrum disorder in adolescents

Rule No.	Association Rules	Accuracy
1	A3_Score=1∩A5_Score=1∩A6_Score=1∩A9_Score=1→ASD=Yes	0.9946
2	A3_Score=1∩A5_Score=1∩A8_Score=1→ASD=Yes	0.9945
3	A7_Score=1∩A10_Score=1→ASD=Yes	0.9944
4	A1_Score=1∩A4_Score=1∩A5_Score=1∩A8_Score=1→ASD=Yes	0.9943
5	A3_Score=1∩A5_Score=1∩A6_Score=1∩A10_Score=1∩gender=Male → ASD=Yes	0.9932
6	A1_Score=1∩A7_Score=1∩gender=Female→ ASD=Yes	0.9925
7	A3_Score=0∩A10_Score=0→ASD=No	0.9930
8	A4_Score=0∩A5_Score=0→ASD=No	0.9917
9	A3_Score=0∩A5_Score=0→ASD=No	0.9912
10	A1_Score=0∩A5_Score=0→ASD=No	0.9906

Note. The symbol "∩" denotes the logical AND (intersection) between multiple conditions in an association rule. (A1- A9) are personal behavioural questions on the ASD screening data. The score of 1 represents the presence of the behaviour, whereas 0 represents the absence of the behaviour

Table 10
Association rules generated by the tertius algorithm for detecting autism spectrum disorder in adolescents

Rule No.	Association Rules	Confirmation
1	A3_Score=1∩A5_Score=1∩A6_Score= 1→ASD=Yes	0.6652
2	A5_Score=1∩A6_Score=1∩A10_Score= 1→ASD=Yes	0.6398
3	A6_Score=0 ∪ A10_Score = 0 ∪ age =13 → ASD=No	0.6909
4	A3_Score=0 ∪ A5_Score=0 ∪ A6_Score= 0 → ASD=No	0.6652
5	A3_Score=0 ∪ A5_Score=0 ∪ country_of_res = Bahrain → ASD=No	0.6568
6	A3_Score=0 ∪ A5_Score=0 ∪ A9_Score=0 → ASD=No	0.6515
7	A3_Score=0 ∪ A5_Score=0 → ASD=No	0.6322

Table 11
Association rules generated by the apriori algorithm for detecting autism spectrum disorder in adults

Rule No.	Association Rules	Confidence
1	A6_Score=0∩A9_Score=0∩jundice=0→ASD=No	0.970
2	A6_Score=0∩A9_Score=0→ASD=No	0.960
3	A6_Score=0∩A9_Score=0∩autism=0 →ASD=No	0.960
4	A9_Score=0∩jundice=0∩autism=0 →ASD=No	0.930
5	A9_Score=0∩jundice=0→ASD=No	0.930
6	A9_Score=0	0.930
7	A9_Score=0∩autism=0 →ASD=No	0.930
8	A6_Score=0∩jundice=0∩autism=0 →ASD=No	0.910
9	A6_Score=0∩autism=0 →ASD=No	0.910
10	A6_Score=0∩jundice=0→ASD=No	0.900

Notably, the algorithm did not generate similar rules for adults with ASD, implying that these patterns were distinct from those of healthy individuals. Overall, the confidence levels ranged from 90.00% to 97.00%.

In Table 12, the predictive apriori algorithm highlighted crucial factors for identifying ASD in adults, achieving an accuracy of 99.44%. These factors included A1, A4, A6, A8, and A9 scores, all being "Yes". In contrast, the absence of A5, A6, and A9 scores predicted non-ASD adults with an accuracy of 99.46%.

Similarly, in Table 13, the Tertius algorithm revealed that the significant factors for identifying ASD in adults (with a confirmation of 57.54%) included A6_score =Yes, or A9_score =Yes, or the country of residence being Malaysia. For healthy adults, the significant factors (with a confirmation of 56.76%) were A6 Score =No and

A9 Score =No. A comparison of the three algorithms revealed that a positive response for detecting listener boredom (A6) and easily reading emotions (A9) may indicate the presence of ASD, while a negative response for grasping implied meanings (A5) and easily reading emotions (A9) could suggest the absence of ASD in adults.

Table 12
Association rules generated by the predictive apriori algorithm for detecting autism spectrum disorder in adults

Rule No.	Association Rules	Accuracy
1	A1_Score=1∩A4_Score=1∩A6_Score=1∩A8_Score=1∩A9_Score=1→ASD=Yes	0.9944
2	A4_Score=1∩A5_Score=1∩A6_Score=1∩A8_Score=1∩A9_Score=1→ASD=Yes	0.9943
3	A1_Score=1∩A5_Score=1∩A6_Score=1∩A7_Score=1∩A9_Score=1→ASD=Yes	0.9943
4	A1_Score=1∩A5_Score=1∩A6_Score=1∩A8_Score=1∩A9_Score=1∩A10_Score=1→ASD=Yes	0.9942
5	A1_Score=1∩A3_Score=1∩A4_Score=1∩A5_Score=1∩A6_Score=1∩A9_Score=1∩A10_Score=1∩jundice=0→ASD=Yes	0.9939
6	A5_Score=0∩A6_Score=0∩A9_Score=0→ASD=No	0.9946
7	A3_Score=0∩A5_Score=0→ASD=No	0.9944
8	A4_Score=0∩A5_Score=0→ASD=No	0.9944
9	A3_Score=0∩A4_Score=0∩A6_Score=0→ASD=No	0.9943
10	A5_Score=0∩A7_Score=0∩A9_Score=0∩jundice=0→ASD=No	0.9940

Table 13
Association rules generated by the tertius algorithm for detecting autism spectrum disorder in adults

Rule No.	Association Rules	Confirmation
1	A6_Score=1∪A9_Score=1 ∪ country_of_res=Malaysia → ASD=Yes	0.5754
2	A6_Score=1∪A9_Score=1 ∪ country_of_res=Italy → ASD=Yes	0.5722
3	A6_Score=1∪A9_Score=1 ∪ country_of_res=Bangladesh → ASD=Yes	0.5708
4	Class/ASD=1==>A6_Score=1∪A9_Score= 1 ∪ country_of _res=Ireland → ASD=Yes	0.5694
5	A6_Score=1 ∪ A9_Score=1 ∪ age = 30 → ASD=Yes	0.5692
6	A6_Score=1 ∪ A9_Score=1 ∪ country_of_res = Canada → ASD=Yes	0.5681
7	A6_Score=1 ∪ A9_Score=1 ∪ country_of_res = Russia → ASD=Yes	0.5681
8	A6_Score=1 ∪ A9_Score=1→ASD=Yes	0.5676
9	A6_Score=0 ∩ A9_Score=0→ASD=No	0.5676

Performance Comparisons

Figure 5 demonstrates a comparative analysis of the outputs produced by three association rule mining algorithms apriori, predictive apriori, and tertius across four distinct age groups. Figure 5(a) displays the confidence values generated by the apriori algorithm, indicating that adolescents exhibit the highest confidence (100.00%) and children the lowest (95.00%). Figure 5(b) shows the accuracy values obtained from the predictive apriori method, with all age groups demonstrating consistently high accuracy, ranging between 099.40% and 99.50%. In contrast, Figure 5(c) illustrates the confirmation values derived from the Tertius method, which revealed greater variability across age groups, spanning from 57.50% in adults to 66.80% in children.

We evaluated the computational efficiency of the algorithms on four distinct ASD datasets (Figure 6). The apriori algorithm demonstrated instantaneous execution (0.0 s) across all age groups. In contrast, the tertius algorithm exhibited significantly higher computational demands, with the longest runtime observed in the adult group (5250s). Furthermore, the computational time varied by developmental stage: the toddler group showed the fastest execution, whereas the adult group required the most substantial processing time.

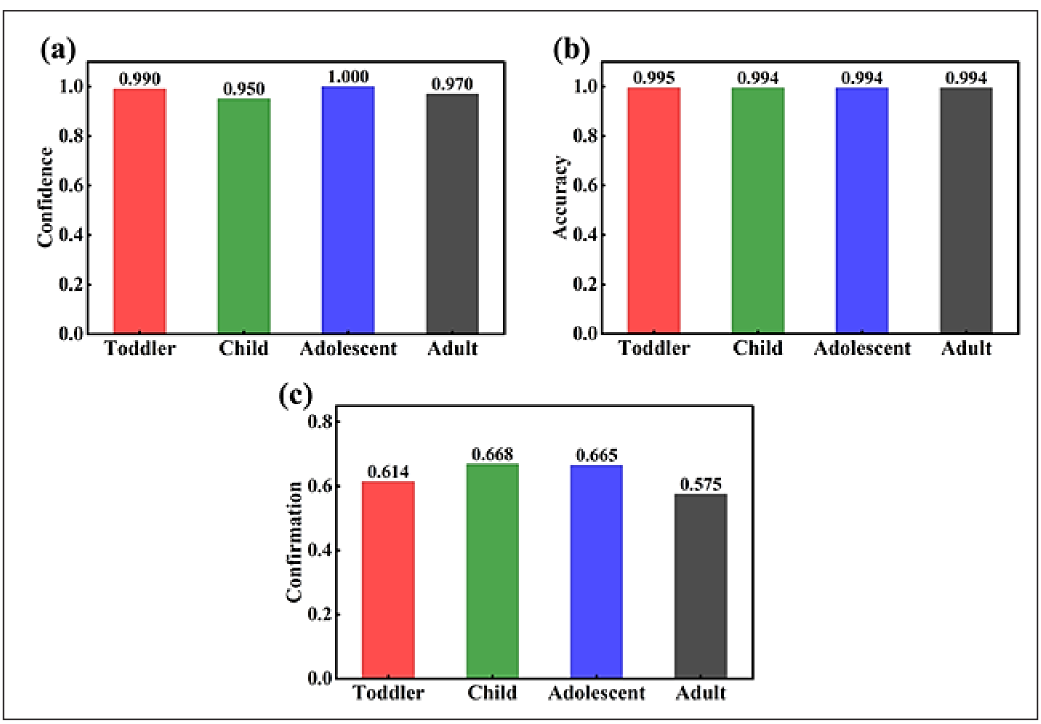

Figure 5. Comparison of the performance of three association rule mining techniques across four developmental age groups: (a) Apriori; (b) Predictive Apriori; and (c) Tertius

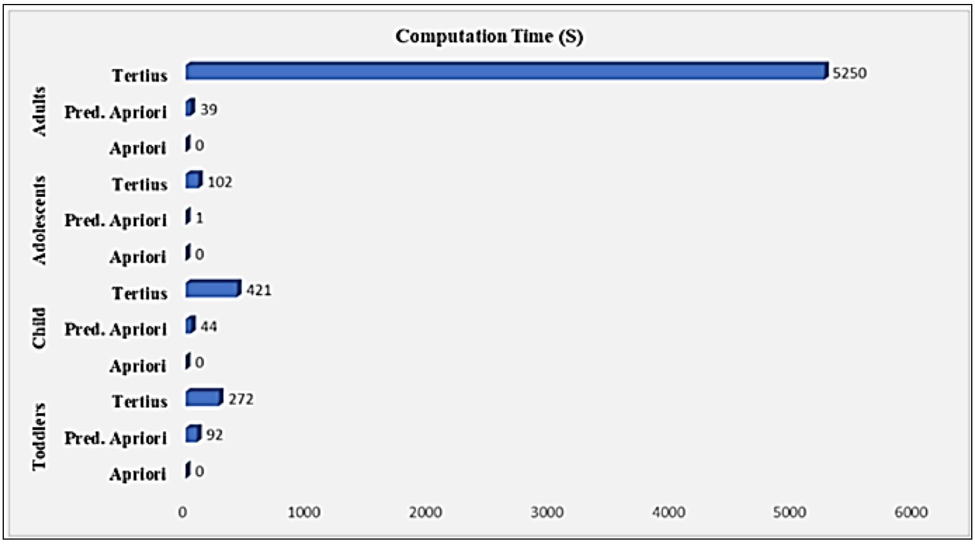

Figure 6. Computational time comparison of different association rule mining algorithms among four age groups

DISCUSSION

In this section, we interpret our findings for each dataset based on the rules found by applying different association rule mining algorithms. Thus, we tried to find some interesting insights into ASD at different ages.

Association Rules and Risk Factors to Detect ASD and Healthy Toddlers

Table 2,3 and 4 show the results of the different rule mining methods on the toddler dataset. We can see that the largest rule for identifying toddlers with ASD was extracted by the predictive apriori algorithm with 99.49% accuracy. The interpretation of this rule shows positive responses for “Sensing Small Sound (A1 Score)”, “Pretending Games (A5 Score)”, “Tracking Vision (A6 Score)”, and “Sensing Others Grief and Comforting (A7 Score)”, indicating a toddler with probable ASD. On the other hand, the apriori algorithm provides the smallest rule (single factor) to identify ASD toddlers with 96.00% accuracy, which indicates that an individual positive response to “Pretending Games (A5 Score)” or “Only Simple Gesturing Capability (A9 Score)” is associated with ASD toddlers. According to the apriori algorithm, a toddler might have ASD, although there is no “family member with PDD”. Again, based on the predictive apriori algorithm, we found that positive responses for “Pretending Games (A5 Score)” and “Only Simple Gesturing Capability (A9 Score)” were associated factors for male toddlers. According to the Tertius algorithm, positive responses for “Pretending Games (A5 Score)” and “Only Simple Gesturing Capability (A9 Score)” are always risk factors associated with “26/16 Months of Age (Age Mons=26/16)” and “Pacifica Ethnicity Group (Ethnicity=Pacifica)”. However, rules for toddlers with no ASD or healthy toddlers indicate that the negative response for both “Pretending Games (A5 Score)” and “Only Simple Gesturing Capability (A9 Score)” is the predominant factor for healthy toddlers’ identification in all algorithms. In addition, a negative response for “Born with Jaundice (Jaundice)” is another contributing factor indicating that toddlers have no ASD. Nonetheless, it was found that the “Country_of_Residence” was not used to identify ASD and healthy toddlers. In addition, it is noticeable that responses to the screening test questions are more important than other features in defining rules to differentiate between ASD and healthy toddlers.

Association Rules and Risk Factors to Detect ASD and Healthy Children

Association rules and risk factors to detect ASD and healthy Children Tables 5-7 show the results of the different rule mining methods on the child dataset. We can see that the largest rule for identifying children with ASD was extracted by the predictive apriori algorithm with 99.44% accuracy. Interpretation of this rule shows a positive response for “Sensing Small Sound (A1 Score)”, “Tracking Conversation (A3 Score)”, “Quick Activity Swapping (A4 Score)”, “Difficulty in Continuing Conversation (A5 Score)”, and “Difficulty in Making Friends (A10 Score)”, indicating a child with probable ASD. Among all the rules found using all the algorithms here to identify children with ASD, the “Quick Activity Swapping (A4 Score)” and “Difficulty in Making Friends (A10 Score)” features were the most common. In the apriori algorithm, “Difficulty in Continuing Conversation (A5 Score)” is also an important feature. In the predictive apriori algorithm, “Tracking Conversation (A3 Score)” with a positive response is another important feature. In addition, features other than the responses to the screening test questions do not contribute to the identification of children with ASD.

However, rules derived from the Tertius algorithm indicate that negative responses for both “Quick Activity Swapping (A4 Score)” and “Difficulty in Making Friends (A10 Score)” are the predominating factors for identifying healthy children regardless of the “Country of Residence (country_of_res)”. Besides that, a negative response to “Sensing Small Sound (A1 Score)”, “Lack of Details Capturing (A2 Score)”, “Tracking Conversation (A3 Score)”, “Difficulty in Continuing Conversation (A5 Score)”, “Social Chit Chat (A6 Score)”, “Difficulty in Tracking Character’s Intention (A7 Score)”, “Pretending Games (A8 Score)”, and “Interpreting Facial Expression (A9 Score)” is also important in identifying healthy children according to the predictive apriori algorithm. A negative response to “Tracking Conversation (A3 Score)” was also found to be important in the tertius algorithm. Again, other features than responses to the screening test questions did not contribute to the identification of healthy children.

Association Rules and Risk Factors to Detect ASD and Healthy Adolescents

Tables 8-10 present the results of various rule mining methods applied to the adolescent dataset. Among the rules generated by the three algorithms to identify adolescents with ASD, the most common traits observed included a positive response to "Tracking Conversation (A3 Score)", "Difficulty in Continuing Conversation (A5 Score)", "Social Chit-Chat (A6 Score)", and "Social Situations (A9 Score)". Notably, the predictive apriori algorithm derived two distinct rules for identifying male and female adolescents with ASD. For males, a positive response to "Tracking Conversation (A3 Score)", "Difficulty in Continuing Conversation (A5 Score)", "Social Chit-Chat (A6 Score)", and "Difficulty in Making Friends (A10 Score)" was the most indicative. In contrast, for females, positive responses to "Finding Patterns (A1 Score)" and "Pretending Games (A7 Score)" were more relevant. The most concise rule for identifying adolescents with ASD, according to the predictive apriori algorithm (with an accuracy of 99.45%), suggests that a positive response to both "Pretending Games (A7 Score)" and "Difficulty in Making Friends (A10 Score)" is sufficient, regardless of sex. Additionally, positive responses to "Quick Activity Swapping (A4 Score)" and "Lack of Empathy (A8 Score)" are also critical features for identifying adolescents with ASD. On the other hand, rules for identifying healthy adolescents indicate that negative responses to both "Tracking Conversation (A3 Score)" and "Difficulty in Continuing Conversation (A5 Score)" are key factors for distinguishing healthy adolescents. Furthermore, for adolescents aged 13 years, negative responses to "Social Chit-Chat (A6 Score)" and "Difficulty in Making Friends (A10 Score)" were found to be indicative of healthy adolescents.

Association Rules and Risk Factors to Detect ASD and Healthy Adults

Association rules and risk factors to detect ASD and healthy adults Tables 11 to 13 show the results of different rule mining methods on the adult dataset. We can see that the largest rule for identifying ASD adults was extracted using the predictive apriori algorithm with 99.39% accuracy. Interpretation of that rule shows the positive response for "Small Sound Sense (A1 Score)", "Multitasking (A3 Score)", "Quick Activity Swapping (A4 Score)", "Read Between Lines (A5 Score)", "Interpreting Facial Expression (A6 Score)", "Empathy (A9 Score)", and "Difficulty in Tracking Character's Intention (A10 Score)" and the negative response of "Born with Jaundice (Jaundice)" indicates an adult with probable ASD. However, the smallest rule derived from the tertius algorithm (with 56.76% confirmation) suggests that the positive response to "Interpreting Facial Expression (A6 Score)" and "Empathy (A9 Score)" can solely identify adults with ASD regardless of their country of residence and age. However, the apriori algorithm had the smallest rule for finding healthy adults (with a confidence level of 93.00%). This shows that an adult does not have ASD if they give a negative answer for "Empathy (A9 Score)". In addition, the negative responses for both "Interpreting Facial Expression (A6 Score)" and "Empathy (A9 Score) were the predominant factors for healthy adult identification in all algorithms. Moreover, a negative response for "Born with Jaundice (Jaundice)" and "Diagnosed with Autism (autism)" is another contributing factor that indicates that adults have no ASD, according to the apriori algorithm. As the same response to this feature is present in both the "ASD=Yes" and "ASD=No" classes, it can be said that "Born with Jaundice (Jaundice)" is not a strong discriminatory factor in defining rules to differentiate adults with ASD and healthy adults.

From the above discussions on all datasets, screening test responses are more helpful than other features in identifying patients with ASD. In addition, those born with jaundice may have no relationship with ASD. It is also notable that male and female ASD patient identification rules might differ. Although the screening test questionnaires were slightly different in different datasets, it is noticeable from the above discussion that there are some similarities among the important features found to identify patients with ASD in all datasets. As the questionnaires were different, we may consider the factors that appeared at least twice in the entire experiment as common risk factors for potential ASD patients regardless of age, sex, and country of residence. These common factors include: "Sensing Small Sounds", "Pretending Games", "Tracking Conversations", "Difficulty in Continuing Conversation", "Difficulty in Making Friends", "Difficulty in Tracking Character's Intentions", and "Quick Activity Swapping". Besides, in case of all datasets, we tried to compute Jaccard index value from the rules that we found among the top N rules extracted by apriori, predictive apriori and tertius algorithms. We have obtained a Jaccard index of 0.0 for all algorithms pair apriori vs predictive apriori, apriori vs tertius, and predictive apriori vs tertius of each dataset in this case which indicates that there is no overlap in the sets of extracted rules when top N-rules are considered. This indicates that every algorithm records diverse behavioural characteristics because of variations in evaluation strategies as well as structural rule constraints.

Conflicts among Rule Sets across Different Algorithms

From the above discussion, we can see that different algorithms that we used for association rule mining find out partially overlapping rule sets, although not fully identical. However, we can also observe some conflicts among the rule sets. For instance, the apriori and predictive apriori algorithms associate Pretending Games (A5 Score) with ASD toddlers, while the tertius algorithm associates not ASD toddlers with co-existing Pretending Games (A5 Score) and Pacifica Ethnicity. Again, apriori algorithm only mines rule sets for ASD=no, while the other two mine rules for both ASD = yes and ASD = no. These conflicts among rule sets across algorithms are shown because of the differences in the evaluation metrices used in different algorithms. apriori algorithm focuses on maximising confidence while predictive apriori maximises expected predictive accuracy. On the other hand, tertius maximises confirmation. So, the conflicts we mentioned can be seen as different approaches of defining interestingness. Where apriori and predictive apriori algorithms find more frequent rules, tertius tries to mine surprising (e.g., high confirmation) rules. For clinical uses we can use some ensemble techniques where we can accept any factor as a high risk of ASD only if it appears in rule sets of at least two algorithms. Thus, we can overcome any biases introduced by a specific algorithm and can make more robust predictions.

Caution in Interpreting Demographic Factors

Although behavioural screening scores are the most important indicators in studying ASD, we still get some association rules with demographic factors like age, country of residence, and ethnicity, especially in the rules derived from the tertius algorithm. Unlike behavioural screening scores, demographic factors may come from uneven sampling and imbalance in the dataset. And so, association rules with demographic factors have quite a high chance to represent correlation rather than causal relation. This correlation can either be spurious or incidental. While interpreting the rules with demographic factors, caution should be taken. Without further controlled validation, these factors should not be interpreted as clinically meaningful.

CONCLUSION

Autism Spectrum Disorder (ASD) remains a critical developmental issue, emphasising the necessity of identifying the key determinants influencing its diagnosis. In our study, we explored the associations between various characteristics and ASD classification by utilising three well-established association rule mining algorithms: Apriori, Predictive Apriori, and Tertius. The apriori algorithm demonstrated the highest confidence of 99.00%, 95.00%, 100.00%, and 97.00% for toddlers, children, adolescents, and adults, respectively. In contrast, predictive apriori exhibited superior accuracy, with peak values of 99.50% (toddler) and a consistently high performance of 99.40% across the child, adolescent, and adult datasets. The tertius algorithm yielded varying confirmation scores: 61.40% for toddlers, 66.80% for children, 66.50% for adolescents, and 57.50% for adults. The association rules uncovered several key behaviours indicative of ASD across developmental stages. For toddlers, "pretending games (A5)" and "tracking vision (A6)" emerged as significant factors, while "quick activity swapping (A4)" and "difficulty in making friends (A10)" were significant for children with ASD. For adolescents, "tracking conversations (A3)" and "difficulty in conversations (A5)" were prominent indicators, and for adults, factors such as "grasping implied meanings (A5)", "detecting listener boredom (A6)", and "easily reading emotions (A9)" were significant in distinguishing individuals with ASD from their healthy counterparts. It is evident that responses to screening tests are more informative than other features for identifying patients with ASD. In addition, a history of jaundice at birth was not associated with ASD. There was also no strong correlation between a familial ASD history and ASD occurrence. Furthermore, we observed that the apriori algorithm required the least manipulation time, followed by the Predictive apriori algorithm, while the Tertius algorithm required the most manipulation time. This research most directly benefits clinicians and diagnosticians by providing data-driven behavioural markers for more accurate and earlier ASD screening across age groups. A short list of red flag behaviours for ASD is the output of this research that can be used by clinicians to assess patients in minutes during check-up. For instance, when screening a patient, if one or more high-risk factors of his/her age group are found in him/her, clinicians can promptly suspect the patient of ASD, and the further evaluation plan would be easier for the clinician. Thus, emphasising these high-risk factors will reduce diagnosis delays. This work is also aiding researchers through a comparative analysis of the effectiveness of rule mining algorithms. Although each of the three algorithms are theoretically complete, later we also evaluated the quality of the extracted rules in terms of some additional evaluation metrics, like lift and reverse confidence to determine the strength of the identified associations. Additionally, future studies will build on this work by examining larger datasets and investigating gender-specific association rules to improve the diagnostic accuracy.

ACKNOWLEDGEMENT

The authors would like to thank Jahangirnagar University and Gono Bishwabidyalay for their support in completing this research. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207-216. https://doi.org/10.1145/170035.170072

Aher, S. B., & Lobo, L. (2012). A comparative study of association rule algorithms for course recommender system in e-learning. *International Journal of Computer Applications*, *39*(1), 48-52. https://doi.org/10.5120/4788-7021

Akter, T., Khan, M. I., Ali, M. H., Satu, M. S., Uddin, M. J., & Moni, M. A. (2021). Improved machine learning based classification model for early autism detection. In *Proceedings of the 2nd International Conference on Robotics, Electrical and Signal Processing Techniques* (pp. 742-747). IEEE. https://doi.org/10.1109/ICREST51555.2021.9331013

Al-Diabat, M. (2018). Fuzzy data mining for autism classification of children. *International Journal of Advanced Computer Science and Applications*, *9*(7), 11-17. https://doi.org/10.14569/IJACSA.2018.090702

Aksu, G., & Dogan, N. (2019). An analysis program used in data mining: WEKA. *Journal of Measurement and Evaluation in Education and Psychology*, *10*(1), 80-95. https://doi.org/10.21031/epod.399832

Cai, Q. R. (2020). Cause analysis of traffic accidents on urban roads based on an improved association rule mining algorithm. *IEEE Access*, *8*, 75607-75615.https://doi.org/10.1109/ACCESS.2020.2988288

Delja, F. (n.d.). *Autism screening data for toddlers* [Data set]. Kaggle.

Engle, K., & Rada, R. (2017). Knowledge-guided mutation in classification rules for autism treatment efficacy. *Health Informatics Journal*, *23*(1), 56-68. https://doi.org/10.1177/1460458215627187

Engle, K. M. (2013). *Data mining and domain knowledge: An exploration of methods to advance medical research* (Doctoral dissertation, University of Maryland, Baltimore County).

Engle, K. M., & Rada, R. (2011). A top-k analysis using multi-level association rule mining for autism treatments. In C. Stephanidis (Ed.), *Universal access in human-computer interaction: Applications and services* (Part IV, Vol. 6768, pp. 328-334). Springer. https://doi.org/10.1007/978-3-642-21657-2_35

Flach, P. A., & Lachiche, N. (2001). Confirmation-guided discovery of first-order rules with tertius. *Machine Learning*, *42*(1-2), 61-95. https://doi.org/10.1023/A:1007656703224

Hamid, Z., Khalique, F., Mahmood, S., Daud, A., Bukhari, A., & Alshemaimri, B. (2024). Healthcare insurance fraud detection using data mining. *BMC Medical Informatics and Decision Making*, *24*(1), Article 112. https://doi.org/10.1186/s12911-024-02512-4

Hareendran, S. A., & Chandra, S. S. V. (2017). Association rule mining in healthcare analytics. In Y. Tan, H. Takagi, & Y. Shi (Eds.), *Data mining and big data: DMBD 2017* (Vol. 10387, pp. 31-39). Springer.https://doi.org/10.1007/978-3-319-61845-6_4

Hasanpour, H., Meibodi, R. G., Navi, K., Shams, J., Asadi, S., & Ahmadiani, A. (2019). Fluvoxamine treatment response prediction in obsessive-compulsive disorder: Association rule mining approach. *Neuropsychiatric Disease and Treatment*, *15*, 895-904. https://doi.org/10.2147/NDT.S200569

Ishibuchi, H., Kuwajima, I., & Nojima, Y. (2007). Prescreening of candidate rules using association rule mining and Pareto-optimality in genetic rule selection. In B. Apolloni, R. J. Howlett, & L. Jain (Eds.), *Knowledge-based intelligent information and engineering systems: KES 2007-WIRN 2007, Part II, Proceedings* (Vol. 4693, pp. 509-516). Springer. https://doi.org/10.1007/978-3-540-74829-8_63

Kane, C. (2018). *Classification using association rules* [Thesis]. Technological University Dublin.

Kaoura, G., Kovas, K., Boutsinas, B., & Hatzilygeroudis, I. (2025). A new hybrid reasoning model based on rules, cases and processes: Application to care of individuals facing autism spectrum disorders. *Knowledge and Information Systems*, *67*(1), 371-401. https://doi.org/10.1007/s10115-024-02228-x

Maitra, S., Akter, N., Zahan Mithila, A., Hossain, T., & Shafiul Alam, M. (2021). Apriori-backed fuzzy unification and statistical inference in feature reduction: An application in prognosis of autism in toddlers. In K. Saeed, N. Gazem, & R. Chaki (Eds.), *Progress in advanced computing and intelligent engineering: Proceedings of ICACIE 2020* (pp. 233-254). Springer. https://doi.org/10.1007/978-981-33-4299-6_20

Mazumdar, P., Arru, G., & Battisti, F. (2021). Early detection of children with autism spectrum disorder based on visual exploration of images. *Signal Processing: Image Communication*, *94*, Article 116184.https://doi.org/10.1016/j.image.2021.116184

Mutter, S., Hall, M., & Frank, E. (2004). Using classification to evaluate the output of confidence-based association rule mining. In G. I. Webb & X. Yu (Eds.), *AI 2004: Advances in artificial intelligence* (Vol. 3339, pp. 538-549). Springer. https://doi.org/10.1007/978-3-540-30549-1_47

Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, *40*(4), 1086-1093. https://doi.org/10.1016/j.eswa.2012.08.028

Niranjan, A., Nitish, A., Shenoy, P. D., & Venugopal, K. R. (2016). Security in data mining: A comprehensive survey. *Global Journal of Computer Science and Technology: C Software & Data Engineering*, *16*(5).

Rubio-Martín, S., García-Ordás, M. T., Bayón-Gutiérrez, M., Prieto-Fernández, N., & Benítez-Andrades, J. A. (2024). Enhancing ASD detection accuracy: A combined approach of machine learning and deep learning models with natural language processing. *Health Information Science and Systems*, *12*(1), Article 20. https://doi.org/10.1007/s13755-024-00281-y

Scheffer, T. (2005). Finding association rules that trade support optimally against confidence. *Intelligent Data Analysis*, *9*(4), 381-395. https://doi.org/10.3233/IDA-2005-9405

Singh, K. N., & Mantri, J. K. (2024). An intelligent recommender system using machine learning association rules and rough set for disease prediction from incomplete symptom set. *Decision Analytics Journal*, *11*, Article 100468. https://doi.org/10.1016/j.dajour.2024.100468

Thabtah, F., & Peebles, D. (2020). A new machine learning model based on induction of rules for autism detection. *Health Informatics Journal*, *26*(1), 264-286.https://doi.org/10.1177/1460458218824711

Towle, P. O., & Patrick, P. A. (2016). Autism spectrum disorder screening instruments for very young children: A systematic review. *Autism Research and Treatment*, *2016*, Article 4624829. https://doi.org/10.1155/2016/4624829

UCI Machine Learning Repository. (n.d.-a). *Autistic spectrum disorder screening data for adolescent dataset* [Data set]. University of California, Irvine.

UCI Machine Learning Repository. (n.d.-b). *Autistic spectrum disorder screening data for children data set* [Data set]. University of California, Irvine.

UCI Machine Learning Repository. (n.d.-c). *Autism screening adult data set* [Data set]. University of California, Irvine.

Uddin, M. J., Ahamad, M. M., Sarker, P. K., Aktar, S., Alotaibi, N., Alyami, S. A., Kabir, M. A., & Moni, M. A. (2023). An integrated statistical and clinically applicable machine learning framework for the detection of autism spectrum disorder. *Computers*, *12*(5), Article 92. https://doi.org/10.3390/computers12050092

Wang, T. H., & Guo, F. (2010). Associating IDS alerts by an improved Apriori algorithm. In F. Yu, X. Peng, H. Liu, J. Shu, & R. Ng (Eds.), *Third International Symposium on Intelligent Information Technology and Security Informatics* (pp. 478-482). IEEE. https://doi.org/10.1109/IITSI.2010.47

Wiratsin, I.-O., & Narupiyakul, L. (2021). Feature selection technique for autism spectrum disorder. In *Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence* (pp. 53-56). ACM. https://doi.org/10.1145/3448218.3448241

Wingate, M., Kirby, R. S., Pettygrove, S., Cunniff, C., Schulz, E., Ghosh, T., Robinson, C., Lee, L. C., Landa, R., Constantino, J., Fitzgerald, R., Zahorodny, W., Daniels, J., Nicholas, J., Charles, J., McMahon, W., Bilder, D., Durkin, M., Baio, J., Christensen, D., Van, K., Braun, N., Clayton, H., Goodman, A., Doernberg, N., & Yeargin-Allsopp, M. (2014). Prevalence of autism spectrum disorder among children aged 8 years-Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010. *MMWR Surveillance Summaries*, *63*(2), 1-21.

Yan, X., Zhang, C., & Zhang, S. (2009). Confidence metrics for association rule mining. *Applied Artificial Intelligence*, *23*(8), 713-737. https://doi.org/10.1080/08839510903208062

Zhao, Q., & Bhowmick, S. S. (2003). *Association rule mining: A survey* (Technical Report No. 2003116). Nanyang Technological University, Centre for Advanced Information Systems.