RESEARCH

Autism screening: an unsupervised machine learning approach

Fadi Thabtah[1*] [ID], Robinson Spencer[2], Neda Abdelhamid[3], Firuz Kamalov[4], Carl Wentzel[2], Yongsheng Ye[2] and Thanu Dayara[2]

Abstract

Early screening of autism spectrum disorders (ASD) is a key area of research in healthcare. Currently artificial intelligence (AI)-driven approaches are used to improve the process of autism diagnosis using computer-aided diagnosis (CAD) systems. One of the issues related to autism diagnosis and screening data is the reliance of the predictions primarily on scores provided by medical screening methods which can be biased depending on how the scores are calculated. We attempt to reduce this bias by assessing the performance of the predictions related to the screening process using a new model that consists of a Self-Organizing Map (SOM) with classification algorithms. The SOM is employed prior to the diagnostic process to derive a new class label using clusters learnt from the independent features; these clusters are related to communication, repetitive traits, and social traits in the input dataset. Then, the new clusters are compared with existing class labels in the dataset to refine and eliminate any inconsistencies. Lastly, the refined dataset is utilised to derive classification systems for autism diagnosis. The new model was evaluated against a real-life autism screening dataset that consists of over 2000 instances of cases and controls. The results based on the refined dataset show that the proposed method achieves significantly higher accuracy, precision, and recall for the classification models derived when compared to models derived from the original dataset.

Keywords: Autism spectrum disorders, Classification, Clustering, Machine learning, Self-organising Map

Introduction

Autism Spectrum Disorder (ASD) is a group of neurodevelopment conditions, which inhibit the brain's natural development [4, 16, 19]. People with ASD exhibit a range of symptoms including impairments in social contact and engagement, sensory disturbances, interest in repetitive activities, and various degrees of intellectual disability [15]. Along with the symptoms exhibited, psychological or neurological conditions are also prevalent in people with ASD, with comparatively predominant hyperactivity and attention deficiency, anxiety, depression, and epilepsy [30]. The specific causes of ASD continue being investigated within the medical community. In recent years, the number of confirmed ASD cases has increased, thus necessitating quicker diagnostic services [46, 47].

Recently, machine learning (ML) technology has become a actively researched alternative for ASD diagnosis as it offers several benefits including shortening the diagnostic time, enhancing the ASD detection rate, and identifying impactful features [19, 20, 45]. ML techniques such as a Self-Organizing Map (SOM) explore the data to identify useful unknown patterns, and then use this information in decision making including for medical screening and diagnosis [6, 27]. An SOM is a clustering approach that categorises the data instances into different groups where instances within the same cluster are similar but different to instances in other groups [13, 26]. The SOM technique utilises an Artificial Neural Network (ANN) to transform data onto a two-dimensional map [31].

One of the challenges in using ML for medical data analysis, including the process of creating a data driven ASD diagnosis, is to create a diagnostic model with stronger predictive power than the screening methods used by clinicians [12, 25]. However, since the screening

---

*Correspondence: fadi@asdtests.com
[1] ASDTests, Auckland, New Zealand
Full list of author information is available at the end of the article

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 2 of 13

class within the dataset is based upon the same screening questions, any automated classification model trained using ML techniques on the data will be biased towards the questions used in the screening methods [52]. This occurs because the decision attribute (class label) is usually assigned to the cases and controls in the dataset using a scoring function derived from behavioural screening methods. For example, Quantitative Checklist for Autism in Toddlers (QCHAT) and Autism Spectrum Quotient (AQ) use scoring functions based on the available questions to produce cut-offs (class labels) to discriminate between individuals undergoing behavioural screening [5, 10, 11]. Since classification algorithms process the dataset to build computer-aided models for ASD diagnosis, we suspect that these models can be biased toward the class label assigned by the behavioural screening method.

To be more specific, when an individual undergoes an autism screening receives a score 6 out of 10 by a screening method such the AQ-Adult-10 questionnaire then according to AQ-Adult-10's scoring function the target class of the screening will be Autistic Trait = Yes (see "Medical screening used" section for more details on scoring functions). This assignment was based on just adding up the points obtained from the answers of the 10 questions rather than a comprehensive assessment by the clinician. Despite that the automatic scoring is beneficial by offering some sort of knowledge it did not consider the questions themselves or their relationships. Therefore, this research investigates the above problem by refining autism screening datasets when class is assigned automatically by the screening method scoring functions using a data driven approach based on unsupervised learning.

The research more specifically investigates SOM approach prior the classification process to derive new unbiased class labels based on the items and their answers in the medical screening questionnaire, and without using any scoring functions. The newly derived class labels, along with the clinician decision, can be used together to reduce any possibility of biased decisions when classification algorithms are employed for ASD screening. This paper investigates the question: Can a new filtering approach based on ML be used to reduce bias in the ASD screening process? The authors aim to build a refined ASD model that minimises the classification biases in autism diagnosis systems. This is achieved in two phases. First, a clustering phase where a SOM is applied to learn new patterns, create a new class label, and to refine the dataset. Second, a classification phase where algorithms are applied to the refined dataset to create ASD classification models which have then been validated on a real dataset of over 2,000 instances. It is

the firm belief of the authors that no models based on ML have utilised a SOM to reduce biased decisions, at least in behavioural applications such as ASD screening and diagnosis.

The dataset for the study was collected from recently developed screening systems called 'ASD Tests' and Autism AI' [37, 46, 47], which are based on the Q-CHAT, AQ-10 Child, AQ-10 Adolescent, and AQ-10 Adult ASD screening questionnaires [4]. The 'ASD Tests' uses a scoring function and 'Autism AI' utilises a classification algorithm based on deep neural network (DNN). A score is generated based on the available cases and controls within the systems' data repositories for a test instance, and a screening result is offered to the diagnostician.

This paper is structure as follows: In "Introduction" section, background information regarding ASD and ML is provided for better understanding of these domains. "Literature review" contains a review of selected published journals and articles which are related to ASD and ML, especially on utilising UL on ASD data. "Methodology" section presents the experimental methodology. "Medical screening used" section presents the ASD dataset used in the study and its features. "Data and features descriptions" section describes the creation and evaluation of the data subset by applying ML techniques and measuring their performance. Finally, in Section VIII the conclusions, limitations, and possible future work are briefly discussed.

Literature review

Stevens et al. [41] applied hierarchical cluster analysis to a sample of 138 school-age children with autism to explore two subgroups, i.e. pre-school age and school age, of Autistic Disorder (AD). The study was carried out with the chosen variables on school-age data including expressive and receptive language measures, nonverbal intelligence (IQ), and normal and abnormal behaviour among many other measures. The results derived from hierarchical analysis for the school-age subgroup were further analysed against pre-school characteristics by repeated measures' analysis of variance (ANOVA) [21]. The findings reflect the presence of two subgroups identified by different levels of social, language, and nonverbal ability, with the higher group displaying cognitive and behavioural scores that are generally average, school-age functioning was strongly predicted by preschool cognitive function. The comparative result of two groups indicated that high IQ is necessary in the presence of severe language impairment, but not adequate for optimal outcome.

Obafemi-Ajayi et al. [33] proposed a hierarchical cluster analysis of the phenotype variables of ASD patients to identify more homogeneous subgroups to aid the

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 3 of 13

Diagnostic and Statistical Manual of Mental Disorders' (DSM)-5 model. The dataset used contains a sample population of 213 patients with ASD derived from the Simons Simplex Collection (SSC) project [36]. The authors chose 24 phenotype variables that spanned seven categories, i.e. ASD-specific symptoms, cognitive and adaptive functioning, language and communication skills, behavioural problems, neurological indicators, and genetic indicators. The result showed that the hierarchical model produced two significantly discrete clusters and one outlier cluster, and as if tree structure proceeds, each predominant subgroup can be further divided and evaluated to unravel more homogeneous and clinically significant groups.

Lombardo et al. [29] applied hierarchical clustering to a dataset of adults: 694 with and 249 without autism, to create five discrete subgroups of autism spectrum conditions (ASC). The authors state that there is a high level of variability in the aetiology, development, cognitive features, behavioural features, and more between individuals diagnosed with ASC. Based on this, the authors wanted to determine if there was a significant statistical basis for new ASC categories. Their approach was limited to the cognitive domain of mentalizing which is measured using items from the Reading the Mind in the Eyes test [10, 11]. Hierarchical clustering produced five significantly discrete clusters within the ASC study population and four clusters in the control group. Of the five ASC clusters, three showed high levels of mentalizing impairment with the remaining two clusters being unimpaired.

Stevens et al. [39] applied cluster analysis to a sample dataset of 2116 instances, i.e. children with ASD, to recognise trends of challenging behaviours observed at home and medical care. The dataset was extracted from SKILLSTM, a proprietary data archive operated by a major regional provider of autism care services (Centers for Disease Control & Prevention, 2017). The authors then used K-means clustering algorithms [28] to extract common behaviour profiles through the specified categories of challenging behaviours, i.e. features. The results of their work indicated that (1) A dominant single challenging behaviour is present in most clusters, (2) Several possible variations were identified in challenging behavioural profiles between male and female populations.

Stevens et al. [40] extended their 2017 work and applied Gaussian Mixture Models [GMM] [34] and hierarchical clustering to a sample of 2400 children with ASD to identify behavioural phenotypes and to improve treatment reaction across those observed phenotypes. The authors found that when receiving treatment, the participants had deficits in eight domains, i.e. language, social, adaptive, cognitive, executive function, academic, play, and motor skills. Therefore, GMM analysis was performed to assess proficiency in one of the eight treatment domains and this revealed 16 subgroups. Further analysis by Hierarchical Clustering found five distinct subgroups, and findings indicated two overlying behavioural phenotypes with unique deficit profiles consisting of subgroups that may differ in severity.

Baadel et al. [8] expanding on their previous work [9] and other work of [49] proposed a new semi-supervised ML method called Clustering-based Autistic Trait Classification (CATC) to be applied to three real datasets (adult, adolescent, and child)—this was obtained by a mobile screening application called 'ASDTest' [44] to enhance the efficiency of detecting ASD traits by reducing data dimensionality and redundancy in the autism dataset. Their approach involved two phases: a clustering phase in which the Outlier-based Multi-Cluster Overlapping K-Means Extension algorithm (OMCOKE) [7] was used, and a classification phase in which Repeated Incremental Pruning to Produce Error Reduction (RIPPER) [16], PART [17], Random Forest [22], Random Trees [22], and Artificial Neural Networks [35] were used to evaluate the performance of the proposed CATC method. The results analysis revealed that CATC offers classifiers with higher predictive accuracy, sensitivity, and specificity by reducing the error rate.

Thabtah et al. [48] developed a filtering method-based feature selection called Variable Analysis (VA) to detect influential features of autism from three datasets collected using a screening application called ASDTests [44],these datasets contain cases and controls related to children, adolescents, and adults. The proposed filtering method was compared with five other filtering methods using supervised learning algorithms including decision trees and rule induction. The results revealed that when used as a feature selection method on autism screening datasets, VA selects highly influential features. To be exact, VA minimized the number of features chosen from the children, adolescent, and adult datasets to just 8, 8, and 6, respectively. These subsets of features, when processed by supervised learning algorithms, derive classification models with good predictive power, at least when using decision trees and rule induction approaches.

Thabtah and Peebles [50] extended the work of Thabtah et al. [49] and proposed a rule induction algorithm called Rules-Machine Learning (RML) to quickly screen autism for individuals. The proposed rule induction algorithm learns simple rules in the form of If–Then using greedy classification in which whenever a rule is generated, all its corresponding data will be discarded in a repetitive process. Experimental evaluation of the RML algorithm was compared with well-known classification algorithms RIPPER, RIDOR, Nnge, Bagging, CART, C4.5, and PRISM against the children, adolescent, and adult

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 4 of 13

datasets using different performance metrics. The results pinpointed that the RML algorithm produces a smaller number of rules when contrasted with other rule induction and decision tree algorithms and maintains the level of performance with respect to accuracy, sensitivity, and specificity. The RML algorithm was also compared with a regression-based model by Thabtah et al. [49] using the same autism screening datasets and showed good predictive performance. Nevertheless, the RML and other conventional ML approaches were sensitive to the class imbalance issue as demonstrated in recent research by Alahmari [2], and Abdelhamid et al. [1].

Tawhid et al. [42, 43] improved the performance of ASD detection based on the electroencephalogram (EEG)—a procedure that measures the brain's electrical activity. Using deep learning technology, the authors developed classification systems that process images of time–frequency spectrogram (EEG signals). Initially, the unprocessed data were transformed and filtered, and then features that were extracted from the images assessed using feature engineering, i.e., principal component analysis (PCA). Lastly, multiple convolutional neural networks were designed and applied to the processed data to build predictive models for ASD prediction. Empirical results show the superiority of the proposed deep learning systems when compared with other conventional machine learning techniques in terms of predictive power and other metrics. Table 1 shows the summary of the related work.

Methodology

Figure 1 below shows the methodology employed in our experiment. The ASD dataset obtained from the 'Autism AI' and 'ASD Tests' mobile application systems [37, 38] is first examined closely in a descriptive analysis provided in "Medical screening used" section. One issue identified

in the dataset was that a new instance is generated every time the test is taken within the app. Therefore, some users may have multiple results recorded if they used the test more than once. Two criteria were used to detect such cases: first using the date feature (which also contains a timestamp) and checking if two consecutive tests are taken within five minutes of each other,second, checking if demographic data (Age, Sex, Ethnicity, Jaundice, FamilyASDHistory, User, and AutismAgeCategory) in the consecutive tests match. If these two criteria are met, the original test instance was kept, and the following test instance removed. This approach is only able to detect if a user takes the test multiple times within five minutes and selects the same options for demographic questions. The same user could of course complete multiple tests at varying times or could try selecting different demographic options. The approach also assumes that if a user completes the test multiple times in succession, their first result is the most accurate, so subsequent results were removed. We filtered out 93 instances, and the remaining instances were kept in a pre-processed dataset (PPDS).

After this pre-processing step, two phases were applied to the PPDS to build an ASD screening model to reduce bias. In the first phase, the SOM clustering algorithm was applied to create a map from the scores of test questions. The SOM algorithm is an adaptation of ANN for unsupervised and competitive learning. It creates a single layer of output neurons (nodes) in a two-dimensional map (grid) with the input vector directly connected to the neurons within the map. The map nodes themselves are not interconnected, which means the values of each map node are hidden from others. The SOM clustering algorithm was applied on the questions Q1–Q10 from the PPDS to create a map that consisted of 225 (15 by 15) nodes. Each node on this map was then assigned a class

Table 1 Summary of the literature review

Study	Sample size	Learning methods used
Literature review summary		
Stevens et al. [41]	138	Hierarchical cluster analysis
Obafemi-Ajayi et al. [33]	213	Hierarchical cluster analysis
Lombardo et al. [29]	943	Hierarchical cluster analysis
Stevens et al. [39]	2116	K-means clustering
Stevens et al. [40]	2400	GMM, hierarchical cluster analysis
Baadel et al. [8]	2048	Multi-Cluster OMCOKE, RIPPER, PART, Random Forest, Random Trees, Artificial Neural Networks
Thabtah et al. [49]	2048	RIPPER, C4.5
Thabtah and Peebles [50]	2048	RML, RIPPER, RIDOR, Nnge, CART, C4.5, PRISM
Tawhid et al. [42, 43]	4657	Convolutional neural networks

Thabtah *et al. Health Information Science and Systems* (2022) 10:26
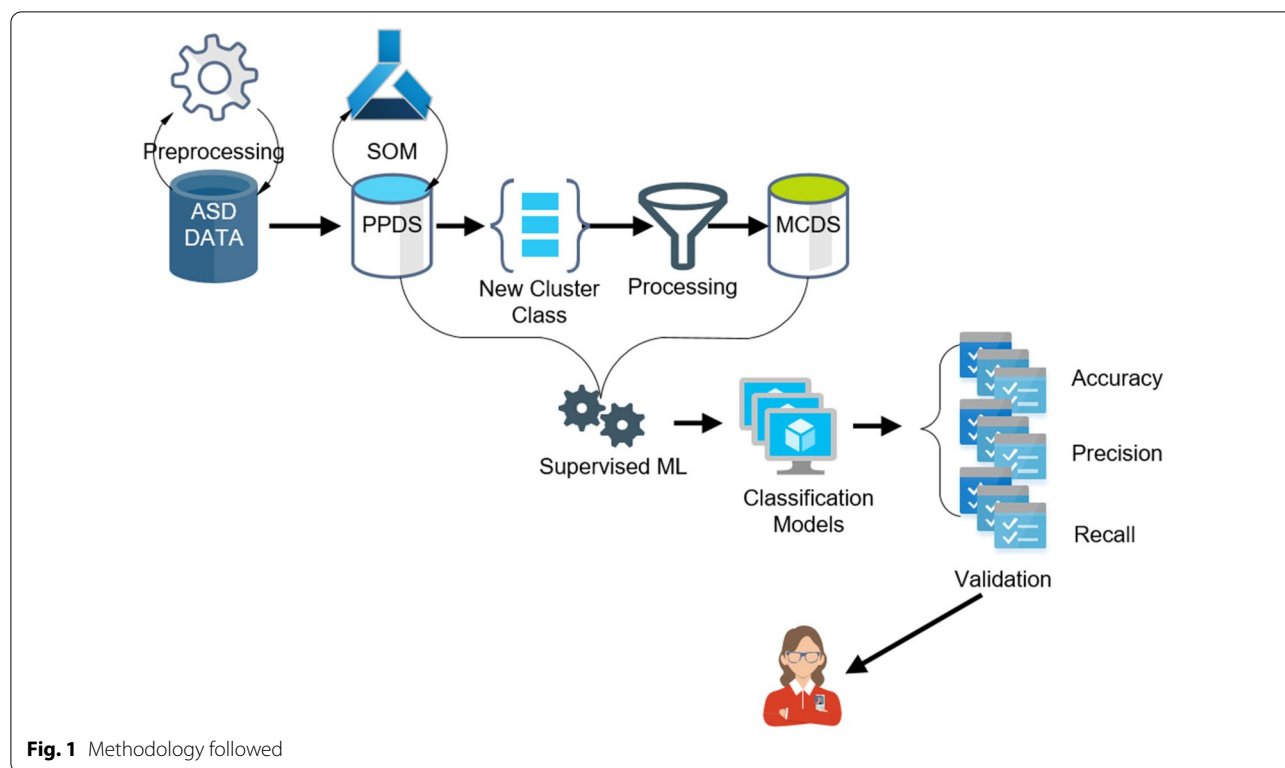
Page 5 of 13

Fig. 1 Methodology followed

value based on the most frequent screening class value for instances mapped onto that node. The application of SOM in this experiment is further detailed in "Data and features descriptions" section.

After creating the new cluster class, it was important to examine the instances where three different screening approaches ('Autism AI'-DNN class, Cluster SOM class, and the Medical Screening Method class) gave the same outcome. By investigating whether data instances where the ('Autism AI'-DNN class, Cluster SOM class, and the Medical Screening Method class) do not match, it may be possible to reduce possible class labels bias in the dataset. To that end, 274 instances were removed from the PPDS where the screening, DNN, and cluster class values did not match to create a new 'Matching class' (MCDS) data subset containing 1681 instances with matching class values. MCDS is the refined dataset where all class labels, including the new cluster label, match with respect to their values for the cases and controls.

In the second phase of the methodology, classification models were trained on the MCDS using two different learning algorithms: Naïve Bayes [18] and Random Forest [14]. These algorithms were selected due to their popularity, ease of implementation, and good predictive performance in multiple previous medical applications [32, 51, 53]. A model's performance is measured by predictive accuracy, precision, and recall. In particular, we compare models trained on the original PPDS dataset with those derived from the refined MCDS dataset using the chosen ML algorithms (See Table 4). This process was performed in two trials: first, building classification models to predict the screening classes, and second, building classification models to predict if the subject has been diagnosed with ASD by a clinician. The building and evaluation of classification models is further outlined in Section V.

The proposed approach is unique and differs from existing work related to improving ASD classification using intelligent techniques such as AI and machine learning. In particular, the proposed approach is one of the first attempts to assess whether the ASD screening decision assigned by the automated medical assessment method using data driven methodology (unsupervised learning) could be biased. The proposed approach improves the screening detection by reducing the class bias by only assessing behavioural features (the medical screening elements) using SOM, and without considering the final score computed from these elements. However, most of the existing research works on ASD classification using behavioural indicators have focused on building classification systems from the dataset in which the class label was primarily or partly assigned using the medical screening score.

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 6 of 13

Medical screening used

Autism Spectrum Quotient (AQ) is a behavioural screening method for autism that is self-administered and can be conducted by adults with an average intelligence level. In its original version, AQ consisted of 50 questions that cover different social and behavioural areas such as attention, social skills, communication, and imagination among others. When the adult goes through the AQ method a score of all questions answered that range between 0 and 50 will be given in which the higher the value of the score the higher autistic traits.

To simplify the autism screening process, and to reduce the screening time of AQ methods Allison et al., [4] proposed different short versions of AQ methods that also can cover besides adult, children, adolescent called AQ-Adult, AQ-Child, and AQ-Adolescent respectively. In these versions, each short questionnaire will contain 10 questions instead of 50 that have been designed to fit the age category of the individual undergoing the screening process. The AQ adolescent questionnaire is intended for individuals with age between 12 and 15 years, and the AQ-Child is a parent-administered questionnaire intended for children between 4 and 11 years. Tables 2, 3, 4, 5 shows the list of questions for each age category. Evaluation of the AQ-Adult, AQ-Child, and AQ-Adolescent autism screening questionnaires have shown similar

Table 2 Features collected and their descriptions for adults based on AQ-10 Adult medical questionnaire [4]

Feature	Type	Question description
Question 1 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	I often notice small sounds when others do not
Question 2 Answer	4 options—after encoding: Binary (0, 1)	I usually concentrate more on the whole picture, rather than the small details
Question 3 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	I find it easy to do more than one thing at once
Question 4 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	If there is an interruption, I can switch back to what I was doing very quickly
Question 5 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	I find it easy to 'read between the lines' when someone is talking to me
Question 6 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	I know how to tell if someone listening to me is getting bored
Question 7 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	When I'm reading a story I find it difficult to work out the characters' intentions
Question 8 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant etc.)
Question 9 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	I find it easy to work out what someone is thinking or feeling just by looking at their face
Question 10 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	I find it difficult to work out people's intentions

Table 3 Features collected and their descriptions for adolescents based on AQ-10 Adolescent medical questionnaire [4]

Feature	Type	Question description
Question 1 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he notices patterns in things all the time
Question 2 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he usually concentrates more on the whole picture, rather than the small details
Question 3 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	n a social group, s/he can easily keep track of several different people's conversations
Question 4 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	If there is an interruption, s/he can switch back to what s/he was doing very quickly
Question 5 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he frequently finds that s/he doesn't know how to keep a conversation going
Question 6 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he is good at social chit-chat
Question 7 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	When s/he was younger, s/he used to enjoy playing games involving pretending with other children
Question 8 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he finds it difficult to imagine what it would be like to be someone else
Question 9 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he finds social situations easy
Question 10 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he finds it hard to make new friends

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 7 of 13

Table 4 Features collected and their descriptions for children based on AQ-10 Child medical questionnaire [4]

Feature	Type	Question description
Question 1 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he often notices small sounds when others do not
Question 2 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he usually concentrates more on the whole picture, rather than the small details
Question 3 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	In a social group, s/he can easily keep track of several different people's conversations
Question 4 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he finds it easy to go back and forth between different activities
Question 5 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he doesn't know how to keep a conversation going with his/her peers
Question 6 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he is good at social chit-chat
Question 7 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	When s/he is read a story, s/he finds it difficult to work out the character's intentions or Feelings
Question 8 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	When s/he was in preschool, s/he used to enjoy playing games involving pretending with other children
Question 9 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he finds it easy to work out what someone is thinking or feeling just by looking at their face
Question 10 Answer	Multiple choice—4 options—after encoding: Binary (0, 1)	S/he finds it hard to make new friends

Table 5 Features collected and their descriptions for toddlers based on Q-Chat-10 medical questionnaire [4]

Feature	Type after encoding	Question description
Question 1 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	Does your child look at you when you call his/her name?
Question 2 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	How easy is it for you to get eye contact with your child?
Question 3 Answer	Multiple choice—45 options—after encoding: Binary (0, 1)	Does your child point to indicate that s/he wants something (e.g. a toy that is out of reach)?
Question 4 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	Does your child point to share interest with you (e.g. pointing at an interesting sight)?
Question 5 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	Does your child pretend (egg care for dolls, talk on a toy phone)?
Question 6 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	Does your child follow where you're looking?
Question 7 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them (e.g. stroking their hair, hugging them)?
Question 8 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	Would you describe your child's first words as: Very typical, Quite typical, Slightly unusual, Very unusual, My child doesn't speak
Question 9 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	Does your child use simple gestures (e.g. wave goodbye)?
Question 10 Answer	Multiple choice—5 options—after encoding: Binary (0, 1)	Does your child stare at nothing with no apparent purpose?

performance in terms of sensitivity and specificity when compared with the original AQ method.

Each question of the short versions of the QA-10 is associated with 4 possible answers for the respondents: 'Definitely Agree,' 'Slightly Agree,' 'Slightly Disagree,' and 'Definitely Disagree.' Conventionally, the screening method will consider a single point per question in AQ-Adult, AQ-Child, and AQ-Adolescent. To be more specific if the respondent's answer to questions 1, 7, 8, and 10 is 'Slightly Agree' or 'Definitely Agree' 1 point will be added. Further, when the respondent's answer to questions 2, 3, 4, 5, 6, and 9 is 'Slightly' or

'Definitely Disagree', a point is added. Then, the points are added up to derive a final score for the respondent. The final score will then be utilised to indicate whether the respondent is associated with autistic traits that is when the final score is above 6. It should be noted that the score computations for each short version of the AQ are different.

Q-CHAT-10 is a shorter version of the Q-CHAT (Baron-Cohen et al., 1992) which is a questionnaire administred by a medical specialist based on a report submitted by the child's parents observing the child's behaviour. Q-CHAT-10 consists of 10 questions with the below sets of possible resposnes:

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 8 of 13

- Set 1: Questions (3,4,5,6,9,10)- Many times a day, A few times a day, A few times a week, Less than once a week, Never
- Set 2: Questions (1,7)- Always, Usually, Sometimes, Rarely, Never
- Set 3:Question 8—Very typical, Quite typical, Slightly unusual, very unusual, My child does not speak

A point of 1 is given when the respondent answers C, D or E for questions 1–9. For question 10, if the respondent answers A, B or C a point is considered. If the total score of the questions yields more than 3 or above then the child is classified as being associated with autistic traits, and he/s will be referred for further assessments.

Data and features descriptions

The dataset in the study was gathered using recently developed screening systems called 'Autism AI' and 'ASD Tests' [38, 45]. It contains 2,048 instances (rows) and 23 attributes (columns). The screening systems contain questionnaires based on the autism screening questionnaires according to Allison et al. [4], namely, Q-CHAT, AQ-10 Child, AQ-10 Adolescent, and AQ-10 Adult. The behavioural attributes Q1–Q10 are based on questionnaires described in Tables 2, 3, 4, 5 respectively. These attributes along with others have been used to form the dataset described in Table 7. We encoded the answers given by the participants in the dataset by assigning '1' if the respondent has answered "slightly agree" or "definitely agree", and '0' (zero) if he/s answered "slightly disagree" or "definitely disagree" during the screening process using the screening systems. The age category for infants, children, adolescents, and adults contains instances for individuals between 0 and 3 years, 4 and 11 years, 12 and 16 years, and above 16 years, respectively. A respondent answers the screening questions using 'Autism AI' or 'ASD Tests' systems, and a score is determined based on the responses they submit with a score between 0 and 10. The 'Class' attribute is then given a YES or a NO depending on the score. The screening systems are designed to deliver YES if the score is above three in infants and above six in all other age categories, otherwise, the class is labelled as NO.

During the data gathering of the screening systems (apps), there was no direct contact with participants; the screening systems offer information to the participants about the use of data in a disclaimer. More importantly, the apps pinpoint that the data collected is strictly for research purposes and the participants consent to the utilization of data during undergoing the screening process. The participants are anonymous as no sensitive information such as names, date of birth, etc. are collected. The

data is public, and the authors have obtained an ethical approval from the University of Huddersfield, United Kingdom [46, 47].

Apart from the screening class label, there are two other class attributes present in the dataset, i.e. 'DNN Prediction' and 'Is ASD Diagnosed'. 'DNN Prediction' is the prediction provided by algorithms based on the historical data of the 'ASD Tests' application, 'Is ASD Diagnosed' is the answer provided by the respondent to the query: 'Has the respondent received a formal ASD diagnosis?' appears at the close of the test. Other attributes include 'ID', 'Date', 'Sex', 'Ethnicity', 'Jaundice' (was the baby born with jaundice?), 'Family ASD History' (has anyone in the immediate family been diagnosed with autism?), 'User' (who is completing the test—categorised as self-test, parent, healthcare professional, etc.). The complete features used in the dataset is shown Table 6.

Classification of the ethnicity of the respondents by 'Class' was carried out to identify the demographics of the participants as well as to learn which region of the world has more participants with ASD traits. The outcome revealed that white Europeans represent slightly more than 50% of data, i.e. 1092 participant records of which 640 identified as having ASD. The reason for more data observations associated with white European is because the screening systems has been used more in countries where respondents were majority white. We do not expect that this may impact the classification results or have any variations since we have good representations in the dataset for other ethnicities including Asian, Black, and Middle Eastern among others; however, we intend in near future to consider evaluating whether some demographic attributes like gender and ethnicity can impact the outcome of the screening.

Figure 2 indicates ASD by age. The size of the dataset varies between the four categories, however, ASD traits are present primarily among infants. It was observed that 142 infants have ASD traits, and all other categories fall below 40 cases for each age group.

Experiments and results analysis

The clustering process using SOM was completed in R Studio development environment [3, 24] on a device with a 2.81 GHz processor and 16 GB of RAM. The SOM train function was called from the SOM.nn library. All default parameter settings were used apart from alpha (training rate) which was set to 0.2, length (training steps) which was set to 10,000, xdim and ydim (map dimensions) which were both set to 15.

Figure 3 shows the map created by the SOM algorithm with blue depicting the share of instances mapped onto each node with screening class 'YES' and red depicting instances with screening class 'NO'. This map is then used

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 9 of 13

Table 6 Feature Description in the dataset

No	Attribute	Data type	Possible values (count)
1	ID	Numeric	Range: 100,078—102,165
2	Date	Date and Time	Range: 05/08/2018–12/02/2020
3	Q1	Binary	1—(1,589) 0—(459)
4	Q2	Binary	1—(938) 0—(1,110)
5	Q3	Binary	1—(826) 0—(1,222)
6	Q4	Binary	1—(780) 0—(1,268)
7	Q5	Binary	1—(983) 0—(1,065)
8	Q6	Binary	1—(869) 0—(1,179)
9	Q7	Binary	1—(1,097) 0—(951)
10	Q8	Binary	1 – (1,342) 0—(706)
11	Q9	Binary	1—(891) 0—(1,157)
12	Q10	Binary	1—(1,409) 0—(639)
13	Age	Integer	Range: 0—80 Mean: 20.3
14	Sex	String	Male—(1,270) Female—(778)
15	Ethnicity	String	White European—(1,092) Asian—(266) Mixed—(108) Middle Eastern—(88) Black—(81) Latino—(70) South Asian—(63) Hispanic—(39) Native Indian—(24) Pacifica—(19) Aboriginal—(8) Others—(190)
16	Jaundice	Boolean	Yes—(373) No—(1,675)
17	Family ASD History	Boolean	Yes—(506) No—(1,542)
18	User	String	Self—(1,231) Parent—(475) Family Member—(149) HealthCare Professional—(42) Others—(151)
19	Autism Age Category	String	Chat—(401) Child—(278) Adolescent—(196) Adult—(1,173)
20	Score	Integer	Range: 0—10 Mean: 6
21	Class	Boolean	Yes—(1,136) No—(912)

Table 6 (continued)

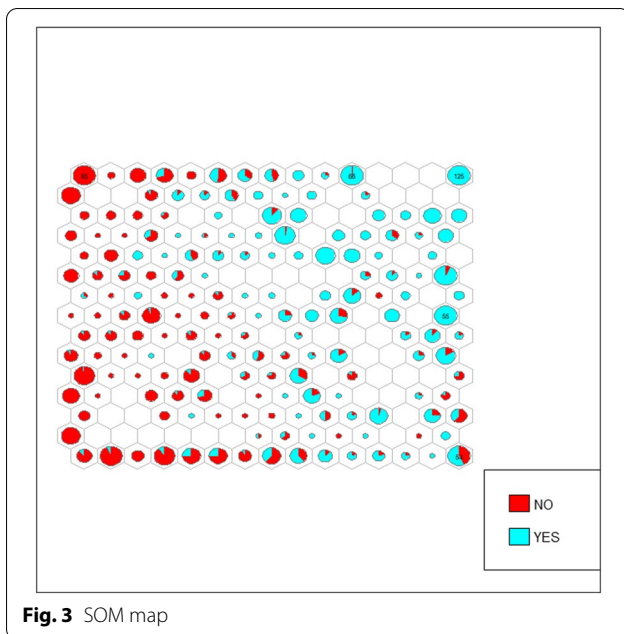| No	Attribute	Data type	Possible values (count)
22	DNN Prediction	Binary	1—(1,138) 0—(910)
23	Is ASD Diagnosed	Binary	1—(715) 0—(1,333)

Fig. 2 ASD screening result by age

to create a new class where each instance mapped to a node is assigned the most frequent screening class value within that node. This new class will be referred to in this paper as the cluster class.

The SOM validate function was used to find the level of similarity between the screening class and the cluster class generated from the SOM algorithm. The cluster class was able to predict the screening class with an accuracy of 85.57%. The resulting predicted values were exported as a new cluster class. The MCDS was then created by filtering out 274 instances where the screening class, DNN class, and the new cluster class did not match.

Table 7 shows the two datasets (PPDS and MCDS) description. The first trial of experiments sought to evaluate the created MCDS by deriving classification models using two algorithms, Naïve Bayes and Random Forest as shown in Table 8. Table 8 includes the results derived by the algorithm in the following scenarios.

1. Dataset = MCDS; Class label = The screening diagnosis assigned by the medical questionnaires
2. Dataset = PPDS; Class label = The screening diagnosis assigned by the medical questionnaires
3. Dataset = PPDS; Class label = The DNN algorithm assigned class

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 10 of 13

Fig. 3 SOM map

4. Dataset = PPDS; Class label = The SMO algorithm cluster class.

Classification and evaluation of the models derived using the ML algorithm were completed within WEKA's Experiment Environment [23] on a device with a 2.81 GHz processor and 16 GB of RAM. Additionally, tenfold cross validation was used to evaluate each model's performance over 10 repetitions. For each dataset

and algorithm combination, mean and standard deviation of accuracy, precision, and recall were recorded. A corrected Paired t-test was used to determine if the difference in performance on the matching class subset was statistically significant (with $P \leq 0.05$).

Table 8 shows the mean accuracy, precision, and recall of classification models built from the PPDS and MCDS dataset under previously mentioned scenarios. Under all metrics, the Random Forest algorithm performed better than the Naïve Bayes algorithm when applied to the same dataset. More importantly for this paper's research question, algorithms applied to the MCDS data subset produced significantly better models (with $P \leq 0.05$) under all considered metrics, compared to models built on the original PPDS data subset. This indicates that instances with matching screening, DNN, and cluster-based classes are easier to classify than instances that do not match. Removing the bias from the original dataset by using the SOM algorithm indeed improved the ASD predictive performance for the ML algorithms considered.

The second trial of ML experiments focuses on predicting the 'Is Diagnosed' feature as a class label (A question in the screening systems/apps) which is reported by the subject undergoing the screening test. In this experiment, we used the same data features in two scenarios.

1. Dataset = MCDS; Class label = The Formal Diagnosis if any reported by the respondent during the screening assessment
2. Dataset = PPDS dataset when the Class label = The screening diagnosis assigned by the medical questionnaires is similar to 'is Diagnosed' feature values
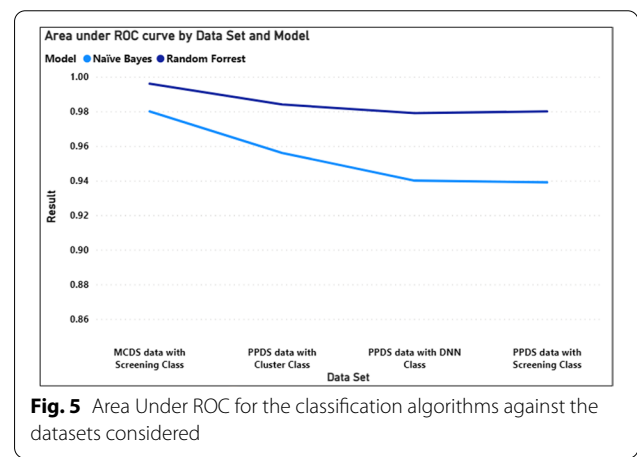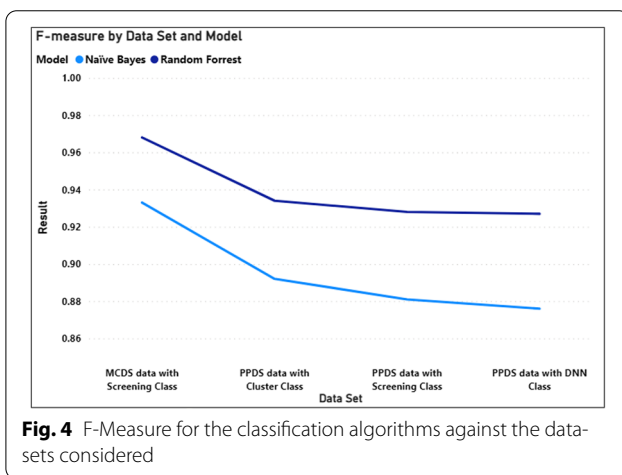
Table 7 PPDS and MCDS datasets description

Dataset	Description	# of Data examples
A pre-processed dataset (PPDS)	The original autism dataset (Table 2) but after pre-processing	1955
Matching class dataset (MCDS)	Subset of PPDS data where the medical screening class, the DNN algorithm class, and the clustering of SOM class values did match	1681

Table 8 First trial results

Algorithm	Metric	MCDS data with screening class (%)	MCDS data with screening class (%)	PPDS data with DNN class (%)	PPDS data cluster class (%)
Naïve Bayes	Accuracy	92.77	86.56	86.31	88.41
	Precision	92.56	86.40	86.19	88.32
	Recall	94.39	89.85	89.56	90.09
Random Forrest	Accuracy	96.15	91.90	91.61	92.79
	Precision	95.65	91.49	91.35	92.06
	Recall	97.39	94.09	93.70	94.60

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 11 of 13

Table 9 Second trial results against the MCDS dataset

Algorithm	Evaluation metric	MCDS dataset—Class label considered in the experiment	
		Class label = The screening method Diagnosis (Yes/No) (%)	Class label = 'Is Diagnosed' feature (Reported by the Respondents during the screening process) (Yes/No) (%)
Naïve Bayes	Accuracy	66.44	65.64
	Precision	74.06	73.51
	Recall	75.98	73.88
Random Forrest	Accuracy	68.28	67.93
	Precision	72.78	72.06
	Recall	83.30	82.89

Fig. 4 F-Measure for the classification algorithms against the datasets considered

Fig. 5 Area Under ROC for the classification algorithms against the datasets considered

We want to assess whether the formal diagnosis class has any bias despite being not validated in the original dataset as reported by the subject undergoing the ASD screening process. Table 9 presents the mean accuracy, precision, and recall of classification models built from the two datasets chosen in the second trial.

The Random Forest algorithm achieved higher accuracy and recall than Naïve Bayes, but a reduced precision rate. Both algorithms when applied to the MCDS data subset produced models with slightly higher accuracy precision and recall than models built using the original PPDS data subset. However, the corrected paired t-test did not consider this difference statistically significant. It shows that the proposed approach also achieved a better performance in the second trial. Nevertheless, the performance rate as expected for the 'Is Diagnosed with ASD' class is below those of the medical screening method class due to several reasons discussed in the limitation section.

We further investigated the area under the receiver operating characteristic (ROC) of the ML algorithms. Figure 4 shows the ROC curve for each model when applying the selected ML algorithms to each data set. The

ROC curve is a commonly used graph representing the trade off between a models true positive rate and false positive rate as its threshold of discrimination changes. Here we are using the area under the ROC curve as a measure of model performance. As with other metrics, classifiers built using the Random Forest algorithm achieved higher area under ROC curve across all datasets. The model built when applying Random Forest to the MCDS data set achieved the highest area under ROC curve with 99.60%.

Figure 4 shows the F-Measure achieved when applying the selected ML algorithms to each data set. The F-Measure is defined as the harmonic mean of a model's precision and recall. The ML algorithms achieved significantly higher F-Measure rates on the MCDS dataset than those derived from the other datasets when the same algorithms were applied. Across each dataset, models built using Random Forest achieved a higher F-measure than models built using Naive Bayes.

Figure 5 shows the area under the receiver operating characteristic (ROC) curve for each model when applying the selected ML algorithms to each data set. The ROC curve is a commonly used graph representing the

Thabtah *et al. Health Information Science and Systems* (2022) 10:26

Page 12 of 13

trade-off between a models true positive rate and false positive rate as its threshold of discrimination changes. Here we are using the area under the ROC curve as a measure of model performance. As with other metrics, classifiers built using the Random Forest algorithm achieved higher area under ROC curve across all data sets. The model built when applying Random Forest to the MCDS data set achieved the highest area under ROC curve.

Conclusions and limitations

This research aims to build an ASD classification model that minimizes the classification biases in autism diagnosis systems using a real-life dataset of over 2000 cases and controls. The proposed approach uses a combination of SOM and classification algorithms and involves two phases. First, the SOM algorithm is utilized to create a new cluster class before comparing it with existing class labels to create a new refined data subset which we call MCDS. Second, classification models were built using ML algorithms on the original data subset (PPDS) and MCDS. The method's ability to reduce bias was evaluated by measuring the performance of the automated classification models derived by the ML algorithms from both the MCDS and the PPDS and using different class labels.

The proposed ML approach was successful in increasing the classifier's accuracy, precision, and recall for the models built on the MCDS. This indicates a reduction in bias in data instances with matching class label values. Models built using Random Forest showed higher performance than models built with Naïve Bayes on both subsets of data (MCDS, PPDS). The model produced by Random Forest based on the MCDS dataset, derived with the proposed approach, achieved accuracy of 96%, precision of 96%, and recall of 97%.

Furthermore, the use of the proposed approach improved the accuracy, precision, and recall of the models trained to predict the 'Is Diagnosed' class in the MCDS data subset; however, this was not found to be a statistically significant difference. Random Forest achieved accuracy of 68%, precision of 73%, and recall of 83% when applied to the MCDS data subset. While this class is interesting for comparing the screening result with a clinician's diagnosis, there are some underlying challenges that cause bias in this class which this paper cannot address. Adding an additional question to the 'ASD Tests' & 'Autism AI' screening systems to ask the user if they have received an ASD diagnosis either positive or negative from a clinician could help to remove some of this bias.

The proposed ML approach can derive classification models using Naïve Bayes and Random Forest that are useful for researchers and healthcare professionals interested in the screening of ASD. However, this paper is limited to the two classification techniques and the instances in the dataset used. Exploring rule-based classification can be seen as a way forward since not only can the model with reduced bias be offered to diagnosticians, but also models contain an easily interpretable chunk of knowledge. In future work, it would be interesting to apply this rule-based classification methodology to ASD screening and diagnosis by expanding the current methodology to work with other classification and clustering algorithms.

It should be noted that the three classes in the original dataset (PPDS)—screening method class, DNN class, and SOM Cluster class—have all been generated by modelling the original dataset using behavioural screening methods, the ANN classification algorithm, and unsupervised clustering SOM, respectively. Therefore, building classification models to predict these class values, even when achieving high levels of predictive performance, is limited to their prospective values. Because of this, the study intends to validate whether the proposed approach can improve the classification model's ability to predict the "Is Diagnosed" class which asks the subject if they have been formally diagnosed with ASD by a diagnostician. When building classification models to predict the "Is Diagnosed" class there are some challenges. Firstly, the value of this class is reported by the subject taking the test—this is not a reliable validation method. Secondly, since the mobile application systems are designed to be a screening test for ASD, there may be some subjects that use the app prior to obtaining any clinical diagnosis.

Author details
[1]ASDTests, Auckland, New Zealand. [2]Digital Technologies, Manukau Institute of Technology, Auckland, New Zealand. [3]Abu Dhabi School of Management, Abu Dhabi, UAE. [4]Canadian University Dubai, Dubai, UAE.

References
1. Abdelhamid N, Padmavathy A, Peebles D, Thabtah F, Goulder-Horobin D. Data imbalance impact on autism pre-diagnosis system: an experimental study. J Inform Knowl Manag. 2020;19(1):2040014. https://doi.org/10.1142/S0219649220400146.
2. Alahmari F. A comparison of resampling techniques for medical data using machine learning. J Inform Knowl Manag. 2020;19(1):2040016. https://doi.org/10.1142/S021964922040016X.
3. Allaire, J. RStudio: Integrated development environment for R 2012. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.651.1157&rep=rep1&type=pdf#page=14
4. Allison C, Auyeung B, Baron-Cohen S. Toward brief "red flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. J Am Acad Child Adolesc Psychiatry. 2012;51(2):202–12. https://doi.org/10.1016/j.jaac.2011.11.003.
5. Allison C, Baron-Cohen S, Wheelwright S, Charman R, Pasco J, Brayne G. The Q-CHAT (Quantitative Checklist for Autism in Toddlers): A normally

Thabtah *et al. Health Information Science and Systems (2022) 10:26*

Page 13 of 13

distributed quantitative measure of autistic traits at 18–24 months of age: Preliminary report. J Autism Dev Disord. 2008;38(8):1414–25. https://doi.org/10.1007/s10803-007-0509-7.

6. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ. A systematic review on supervised and unsupervised machine learning algorithms for data science. Superv Unsupervised Learn Data Sci. 2020. https://doi.org/10.1007/978-3-030-22475-2_1.

7. Baadel, S. *A machine learning clustering technique for autism screening and other Applications.* [Doctoral thesis]. 2019; University of Huddersfield.

8. Baadel S, Thabtah F, Lu J. A clustering approach for autistic trait classification. Inform Health Soc Care. 2020. https://doi.org/10.1080/17538157.2019.1687482,1-18.

9. Baadel S, Thabtah F, Lu J *Overlapping clustering algorithms: A review*. Computing Conference (SAI) 2016. London, UK: IEEE.

10. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. J Child Psychol Psychiatry All Discipl. 2001;42(2):241–51.

11. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists, and mathematicians. J Autism Dev Disord. 2001;31(1):5–17. https://doi.org/10.1023/a:1005653411471.

12 Bone D, Bishop SL, Black MP, Goodwin MS, Lord C, Narayanan SS. Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. J Child Psychol Psychiatry. 2016;57(8):927–37. https://doi.org/10.1111/jcpp.12559.

13. Bratchell N. Cluster analysis. Chemom Intell Lab Syst. 1989;6(2):105–25. https://doi.org/10.1016/0169-7439(87)80054-0.

14. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

15. Centers for Disease Control and Prevention (CDS). 2017. *Identified prevalence of autism spectrum disorder*. http://www.cdc.gov/ncbddd/autism/data.html

16 Crane L, Batty R, Adeyinka H, Goddard L, Henry LA, Hill EL. Autism diagnosis in the United Kingdom: Perspectives of autistic adults, parents, and professionals. J Autism Dev Disord. 2018;48(11):3761–72. https://doi.org/10.1007/s10803-018-3639-1.

17 Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioural distinction of autism and ADHD. Transl Psychiatry. 2016;6(2):732.

18. Duda RO, Hart PE. Pattern classification and scene analysis. New York: Wiley; 1973.

19. Elsabbagh M, Divan G, Koh Y-J, Kim YS, Kauchali S, Marcín C, Montiel-Nava C, Patel V, Paula CS, Wang C, Yasamy MT, Fombonne E. Global prevalence of autism and other pervasive developmental disorders. Autism Res. 2012;5:160–79. https://doi.org/10.1002/aur.239.

20. Georgescu AL, Koehler JC, Weiske J, Vogeley K, Koutsouleris N, Falter-Wagner C. Machine learning to study social interaction difficulties in ASD. Front Robot AI. 2019;6:132. https://doi.org/10.3389/frobt.2019.00132.

21. Hester YC An analysis of the use and misuse of ANOVA. 2001.

22. Ho TK *Random Forest - Document analysis and recognition*. Proceedings of the Third International Conference, 1, (pp. 278–282) 1995.

23. Holmes G, Donkin A, Witten IH. *Weka: A machine learning workbench*. In Proceedings of ANZIIS'94-Australian New Zealand Intelligent Information Systems Conference (pp. 357–361). IEEE 1994.

24. Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat. 1996;5(3):299–314.

25. Kazeminejad A, Sotero RC. Topological properties of resting-state fMRI functional networks improve machine learning-based autism classification. Front Neurosci. 2019;12:1018.

26. Kohonen T. The self-organizing map. Proc IEEE. 1990;78(9):1464–80. https://doi.org/10.1109/5.58325.

27. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015;16(6):321–32.

28. Lloyd. K-means clustering—least squares quantization in PCM. IEEE Trans Inf Theory. 1957;28:129–37.

29. Lombardo MV, Lai MC, Auyeung B, Holt RJ, Allison C, Smith P, Chakrabarti B, Ruigrok AN, Suckling J, Bullmore ET, Bailey AJ. Unsupervised data-driven stratification of mentalizing heterogeneity in autism. Sci Rep. 2016;6:1–15.

30. Lord C, Brugha T, Charman T, Cusack J, Dumas G, Frazier T, Jones E, Jones R, Pickles A, State M, Taylor J, Veenstra-VanderWeele J. Autism spectrum disorder. Nat Rev Dis Primers. 2020;6:5. https://doi.org/10.1038/s41572-019-0138-4.

31. Miljkovic D *Brief review of self-organizing maps*. 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) 2017. Doi:https://doi.org/10.23919/mipro.2017.7973581.

32. Moore PJ, Lyons TJ, Gallacher J. Random forest prediction of Alzheimer's disease using pairwise selection from time series data. PLoS ONE. 2019;14(2):0211558. https://doi.org/10.1371/journal.pone.0211558.

33. Obafemi-Ajayi T, Lam D, Takahashi TN, Kanne S, Wunsch D. *Sorting the phenotypic heterogeneity of autism spectrum disorders: A hierarchical clustering model*. In 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (pp. 1–7). IEEE 2015.

34. Reynolds DA, Quatieri TF, Dunn RB. Speaker verification using adapted Gaussian mixture models. Digital Signal Process. 2000;10(1–3):19–41.

35. Rosenblatt. *The perceptron, a perceiving and recognizing automaton*. Project Para, Cornell Aeronautical Laboratory. 1957

36. SFARI. *Simons Foundation Autism Research Initiative*. https://www.sfari.org/resource/simons-simplex-collection/ 2015.

37. Shahamiri SR, Thabtah F. Autism AI: a new autism screening system based on artificial intelligence. Cogn Comput. 2020. https://doi.org/10.1007/s12559-020-09743-3.

38. Shahamiri SR, Thabtah F. *Google Play*. Autism AI: https://play.google.com/store/apps/details?id=com.rezanet.intelligentasdscreener&hl=en 2019

39. Stevens E, Atchison A, Stevens L, Hong E, Granpeesheh D, Dixon D, Linstead E. *A cluster analysis of challenging behaviors in autism spectrum disorder*. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 661–666). IEEE 2017.

40. Stevens E, Dixon DR, Novack MN, Granpeesheh D, Smith T, Linstead E. Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning. Int J Med Inform. 2019;129:29–36.

41. Stevens MC, Fein DA, Dunn M, Allen D, Waterhouse LH, Feinstein C, Rapin I. Subgroups of children with autism by cluster analysis: a longitudinal examination. J Am Acad Child Adolesc Psychiatry. 2000;39(3):346–52.

42. Tawhid MNA, Siuly S, Wang H, Whittaker F, Wang K, Zhang Y. A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from EEG. PLoS ONE. 2021;16(6): e0253094. https://doi.org/10.1371/journal.pone.0253094.

43. Tawhid MNA, Siuly S, Wang H. Diagnosis of autism spectrum disorder from EEG using a time-frequency spectrogram image-based approach. Electron Lett. 2020;56(25):1372–5.

44. Thabtah F. ASDTest: A mobile app for ASD screening. 2017 www.asdtests.com.

45. Thabtah F. Machine learning in autistic spectrum disorder behavioral research: a review and ways forward. Inform Health Soc Care. 2018;44:278–97. https://doi.org/10.1080/17538157.2017.1399132.

46. Thabtah F. An accessible and efficient autism screening method for behavioural data and predictive analyses. Health Inform J. 2019. https://doi.org/10.1177/1460458218796636.

47. Thabtah F, Abdelhamid N, Peebles D. A machine learning autism classification based on logistic regression analysis. Health Inform Sci Syst. 2019;7(1):12. https://doi.org/10.1007/s13755-019-0073-5.

48 Thabtah F, Hammoud S, Kamalov F, Gonsalves A. Data imbalance in classification: Experimental evaluation. Inform Sci J. 2020;513:429–41.

49. Thabtah F, Kamalov F, Rajab K. A new computational intelligence approach to detect autistic features for autism screening. Int J Med Infrom. 2018;117:112–24. https://doi.org/10.1016/j.ijmedinf.2018.06.009.

50 Thabtah F, Peebles D. A new machine learning model based on induction of rules for autism detection. Health Inform J. 2019. https://doi.org/10.1177/1460458218824711.

51. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019;19(1):1–16. https://doi.org/10.1186/s12911-019-1004-8.

52. Vaishali R, Sasikala R. A machine learning based approach to classify autism with optimum behaviour sets. Int J Eng Technol. 2018;7(4):18. https://doi.org/10.14419/ijet.v7i3.18.14907.

53. Wei W, Visweswaran S, Cooper GF. The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data. J Am Med Inform Assoc JAMIA. 2011;18(4):370–5. https://doi.org/10.1136/amiajnl-2011-000101.