



Autism AI: a New Autism Screening System Based on Artificial Intelligence

Seyed Reza Shahamiri¹ · Fadi Thabtah²

Received: 11 September 2019 / Accepted: 10 June 2020 / Published online: 20 June 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Autistic spectrum disorder (ASD) is a neurodevelopment condition normally linked with substantial healthcare costs and time-consuming assessments where early detection of ASD traits can help limit the development of the condition. The existing conventional ASD screening methods contain a large number of items and are based on domain expert rules which may be criticized of being lengthy and subjective. More importantly, these methods use basic scoring functions to pinpoint to autistic traits rather intelligently learning patterns from cases and controls which can be more accurate and efficient. One promising solution to deal with the above issues and speed up ASD assessment referrals is to develop intelligent artificial intelligence screening methods that not only provide accurate pre-diagnostic classifications but also improve the efficiency and accessibility of the screening process. This paper proposes a new autism screening system that replaces the conventional scoring functions in classic screening methods with deep learning algorithms. The system is composed of a mobile application that provides the user interface capturing questionnaire data; an intelligent ASD detection web service that interfaces with a Convolutional Neural Network (CNN) trained with historical ASD cases; and a database that enables the CNN to learn new knowledge from future users of the system. The CNN classification method was evaluated against a large autism dataset consisting of adult, adolescent, child, and toddler cases and controls. The results obtained from the CNN were compared with other intelligent algorithms in which superior performance was achieved by the CNN. Particularly, the proposed CNN-based ASD classification system revealed higher accuracy, sensitivity, and specificity when compared with conventional screening methods. This indeed will be of high benefit for busy medical clinics and diagnosticians and could possibly be a new direction to change the way ASD diagnosis process is conducted in the future.

Keywords Artificial intelligence; · Autism spectrum disorder (SD); · Cognitive computing; · Deep learning; · Intelligent systems; · Machine learning; · Medical screening

Introduction

Autism spectrum disorder (ASD) is a life-long developmental disorder affecting social and communication skills of individuals and impacts the life of their family members [1]. The symptoms of autism are more visible and easier to identify

in children 2 to 3 years of age. According to [2], one out of every 68 children has autism in the USA. Consequently, various diagnostic methods have been developed to identify autistic traits in its primitive stage to readily provide the necessary health support and services.

The diagnosis for ASD can be difficult since there are no typical medical tests, like a blood test, to diagnose ASD. To start the diagnosis process, general practitioners (GPs) often screen cases for the possibility of autistic traits and then refer potentially positive cases to specialized psychologists or psychiatrists for further behavioral evaluation. The ASD diagnosis can be initiated on toddlers aging 18 months or older, although receiving a final diagnosis may occur at later age [3]. The ASD diagnostic process requires medical professionals to conduct a clinical assessment of the individual's developmental age based on a variety of categories (e.g., behavior excesses, communication, self-care, and social skills).

✉ Seyed Reza Shahamiri
admin@rezanet.com

Fadi Thabtah
Fadi.fayez@manukau.ac.nz

¹ Department of Electrical, Computer, and Software Engineering, Faculty of Engineering, The University of Auckland, Auckland, New Zealand

² Manukau Institute of Technology, Auckland, New Zealand

This widely accepted approach is referred to as clinical judgment. There are common diagnostic instruments for ASD such as Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R) in which multiple questions and activities may be evaluated by the diagnosticians during the diagnostic process. In addition, an early diagnosis is also important for neuroplasticity, which is larger when children are younger.

Despite the acceptable levels of accuracy and validity of clinical diagnosis methods, they have been criticized of being time-consuming with respect to assessment time, having massive numbers of questions, using a static basic scoring function for generating the autism score, and needing specialized clinicians to administer the process among others [4, 5]. Thus, scholars at the research areas of applied behavior science and psychology developed screening methods based on the clinical diagnosis methods such as Modified Checklist for Autism in Toddlers (M-CHAT) and Autism Quotient (AQ) to potentially decrease the waiting time for potential individuals on the spectrum and their family members [6]. However, according to [7]–[10], since most screening methods are built on clinical diagnosis methods, they inherit many of their deficiencies such as having a large number of items in the questionnaire and poor accessibility for non-specially trained professionals.

One promising approach to deal with the above issues and speed up ASD assessments referrals is to develop an intelligent screening method that not only provide accurate pre-diagnostic classification but also improve the efficiency and the accessibility of the screening process. The intelligent screening method will utilize state-of-the-art artificial intelligence (AI) techniques to develop a classification model that can predict autistic traits using historical cases and controls rather handcrafted rules with a scoring function. This model will be robust as whenever the input dataset gets updated, the model structure will be amended without any human intervention. In this respect, deep learning and neural network algorithms offer such mechanisms and have proven to be highly effective in different classification problems where traditional ML methods failed to provide accurate models [11, 12]. Using deep learning algorithms, the screening method will be able to make predictions by learning the hidden knowledge and patterns associated with autism by looking at historical data samples.

This paper reports on an accurate autism screening system that replaces the conventional scoring functions in classic screening methods. We propose a deep learning-based screening system, called Autism AI System, that purely utilizes a new screening algorithm based on Convolutional Neural Networks (CNNs) to improve the accuracy of the screening process besides minimizing subjective decisions. We have chosen different machine learning (ML) methods to be the base methods for testing the CNN. Autism AI System is currently accessible via a mobile platform making

it available to a wide variety of stakeholders including patients, family members, caregivers, diagnosticians, teachers, and health professionals, among others. It uses a CNN for predicting autistic traits, hence replacing the scoring function and the handcrafted rules of classic screening tools; this makes ASD traits detection based on actual learning patterns and unbiased. Therefore, Autism AI can speed up the ASD pre-diagnosis process which can indeed help avoiding unnecessary delays for proper healthcare service access (speech therapies, special education, etc.) and minimize the risk of developing further social and communication difficulties. In addition to explaining the proposed Autism AI System, this article investigates the following research questions:

1. Is the deep learning mechanism capable of replacing the conventional scoring functions and static rules to predict autistic traits?
2. And if so, does the deep learning screening algorithm improve the performance of pre-diagnosis of autism in terms of accuracy, sensitivity, and specificity when compared with machine learning screening?

This paper is structured as follows: the “[Literature Review](#)” section introduces the conventional ASD screening methods and then reviews relevant intelligent studies on the utilization of machine learning in detecting autism. This section also compares existing autism screening mobile applications. In the next section we discuss the proposed system and its primary components. The “[Result Analysis and Evaluation](#)” section is devoted to the testing and validation of the CNN prediction algorithm using real autism dataset and several common machine learning techniques. The last section concludes the paper.

Literature Review

This section initially reviews two sets of questionnaire-based screening methods and then surveys the related autism detection ML studies. Finally, a survey on the existing autism screening applications that use a mobile platform is provided.

Conventional ASD Screening Methods

CHAT was developed by [13] as a quantitative checklist for toddlers to be administered by clinicians in which a report is submitted by the child’s parents based on observations of the child’s behavior. A modified version, which enhances CHAT’s low sensitivity, was developed by [14]. Known as Q-CHAT-10, M-CHAT was later shortened to ten questions in order to make it less time-consuming [15]. CHAT-23, the Chinese version of Q-CHAT, extended the screening population to toddlers aged from 16 to 30 months [16]. For each item

in Q-CHAT-10, the respondent has to choose one alternative of Always, Usually, Sometimes, Rarely, and Never. Once all the ten items are answered, a score of “1” or “0” will be assigned to each question. The scoring function then adds up all the 1 s, and when the total score is greater than 3, then there is potential autistic trait and the recommendation will be to take the toddler for further assessment.

For identifying autistic traits in older individuals with an average level of intelligence, AQ was proposed [17]. The AQ questionnaire consists of 50 different questions covering the areas of social skills, attention switching, imagination, communication, and attention to detail. The AQ test has four possible rating responses (“Definitely Agree,” “Slightly Agree,” “Slightly Disagree,” and “Definitely Disagree”) depending on which final score is calculated. The final score can range from 0 to 50, and a higher score indicates an increased level of autistic symptoms. Later, a different version of AQ was launched to cover adolescents and children [18].

To make it simpler and less time-consuming, Alison et al. [15] presented a compressed version of the original AQ adult version known as AQ-Adult-10. The questions of AQ-10 have four possible responses similar to the original AQ. The screening rule often considers one point per question. The overall score is then calculated using a diagnosis rule, and anyone who scores above the threshold of six is suspected to have autism and other related impairments. Lastly, [18, 19] have developed full AQ versions for adolescents and children, respectively, while [15] proposed shorter versions for the full adolescent and children’s AQ tests. Overall sensitivity and specificity of AQ were reported as 77% and 74%, respectively, with a cutoff score of 32 [19].

Related Machine Learning Studies of Autism Detection

Duda et al. [7] investigated the potential use of outcomes based on ML algorithms to assist clinicians conduct ADOS-R (Module 1) diagnosis method. They claimed, based on the results obtained by using different ML techniques, that ADOS-R (Module 1) items can be replaced with just 8 items (common features found in the ML classifiers) from 10 activities and 29 items. Therefore, the efficiency of conducting ADOS-R (Module 1) can be significantly improved. However, later research by [8, 9] revealed serious pitfall in the methodology and implementation of the study conducted by Duda and his colleagues. In particular, it was shown that the study did not consider integrating ML within ADOS-R diagnosis methods rather the authors just applied a number of machine learning algorithms on a static dataset related to autism in a conventional way. Thus, if the dataset characteristics change, the results will indeed change and therefore, such results cannot be generalized.

Another study investigated efficient ways to differentiate between ADHD and obstructive sleep apnea (OSA) [20]. The authors utilized 217 children who had been classified by physicians as having ADHD, OSA, and a mixture of ADHD and OSA according to DSM IV standards. The data were collected using different diagnostic tools. Three ML algorithms were adopted to derive classifiers that can assist clinicians and physicians in improving the diagnostic decision. Reported results indicate that 17 features show substantial difference among three classes of pervasive developmental disorders (PDDs) particularly in the Child Behavior Checklist (CBCL).

With respect to PDD, Wolfers et al. [21] discussed related issues including small sample sizes, external validity, and ML algorithmic challenges without a clear focus on ASD. A review on the applicability of different algorithms such as neural network and decision tree models (Random Forest) to reduce the time of ASD diagnostic process was conducted by [22], while [23] investigated Random Forest algorithm on an autism dataset from Georgia Autism and Developmental Disabilities Monitoring (ADDMM) Network utilizing phrases and words obtained in children’s developmental evaluations. The dataset consists of 5396 evaluations for 1162 children of whom 601 are on the spectrum. The Random Forest classifiers were evaluated on an independent test dataset that contains 9811 evaluations of 1450 children. The results reported that Random Forest achieved around 89% predictive value and 84% sensitivity.

Thabtah [3] critically analyzed pitfalls associated with experimental studies that adopted ML for ASD classification by pinpointing issues related to datasets and learning algorithm methodologies used. Among the issues identified were interpreting the classifiers’ content derived by the learning algorithm, noise in autism datasets, feature selection process, missing values, and class imbalance and embedding the classification algorithm within an existing screening method. Later on, he proposed a new feature selection method to identify influential autistic traits of children, adolescent, and adults [24]. It was reported that five influential features, when processed, are able to show high predictive rate in detecting autism.

Cognitive functions and their correlations with medical diagnosis have been investigated using Bayesian inference systems [25]. The authors have evaluated the effectiveness of using probabilities associated with patients’ symptoms to detect the accuracy of medical illness based on Bayes’ theorem. The authors used data related to liver function tests and patient’s characteristics. The diagnosis in the dataset was primarily based on tissue tests (autopsy, biopsy), and the symptoms considered were selected from the available patients’ records. Computer programs were coded after probabilistically modeling the data in order to obtain a ranked set of illness using calculated symptoms’ probabilities. In this context, we utilized Bayes Net classifier, which employs Bayes’ theorem to estimate prior probabilities linked with the items (questions in

the medical tests) in order to forecast individuals being on the spectrum.

Existing Mobile Screening Applications for Autism

- *ASDTests*: ASDTests [26] is a multilingual mobile application for detecting autistic traits in different age categories. This app is based on different conventional screening methods including AQ-Adult-10, AQ-Adolescent-10, AQ-Child-10, and Q-CHAT-10. The app adopts the classic scoring functions of the aforementioned screening methods, and based on the final score, it outputs the decision of whether a case is associated with autistic traits. The objective of ASDTests was to collect data related to autism in order to utilize it later to improve the performance of the screening process. ASDTests do not use any intelligent machine learning algorithm in predicting the possibility of autistic traits rather it uses conventional rules and scoring functions developed by conventional screening methods. The app comes in ten different languages and it has high rating and 214 reviews on both Google Play Store and Apple App Store. The number of downloads is over 4000 installs at the time of this writing.
 - *Awesomely Autistic*: This is a mobile application that contains multiple choice questionnaire for screening autism [26]. The main aim of the application is to develop a simple autism screening mechanism. This application can be used by general practitioners for referring patients as an evaluator of autism using the conventional AQ screening questionnaire. The Awesomely Autistic app is available in multiple languages as well. It has a 4.1 rating and 54 reviews and has been downloaded more than 10,000 times.
 - *Autism Test*: Autism Test is another mobile application to detect autism for all age groups [27]. Using a screening method, this app consists of 20 questions related to feelings and tasks, but it is unknown what mechanism is applied to design the questions. The outcome of the test is the autism likelihood level that the subject has developed.
- This application has 3.3 ratings and 132 reviews. This screening app has a translated version in Arabic language called Autism Test Light with a 2.5 user rating as well.
- *Autism and Developmental Disorder Screening (ANDDS) app*: This screening app's aim is to evaluate whether children under the age of 36 months and infants 6 months of age or older exhibit autistic traits. The parents or clinicians can use this app by answering a series of yes/no questions covering the child's behavioral aspects at different phases of their age (i.e., 6, 12, 15, 18, 24, and 36 months). Then, the app offers the outcome of the screening in three different colored bands based on the given answers in which green indicates that the child is normal in terms of behavioral development, yellow shows that parents should watch the progress of their child's behavioral development, and red pinpoints that further clinical evaluation is needed for the child. The ANDDS has no rating.
 - *ASDetect*: ASDetect is an autism screening application for infants aged between 11 and 30 months [28]. The development of the application focuses on creating positive outcome on children with ASD. ASDetect contains three categories of screening based on the child age (12, 18, and 24 months). The authors of the app claimed that the screening accuracy reached 81% when compared with conventional screening, but this has not been verified by independent studies yet. The app has been able to gain 4.5 ratings with 39 reviews so far.

Table 1 compares the aforementioned ASD screening mobile applications and Autism AI. As can be seen, there are only few mobile-friendly ASD screening applications available on Google Play Store and Apple App Store that are based on scientifically studied screening methods. More importantly, there is no screening app at all that adopts AI or machine learning techniques to predict autistic traits rather all these screening apps utilize basic scoring functions adopted from the conventional screening methods (questionnaires). For example, Awesomely Autistic and ASDTests both adopt AQ scoring functions to come up with the final score, and based

Table 1 Autism screening applications summary

App name	Developer	Target	Method	Rating	Reviews	Apple iTunes	Google Store	No. of questions	Downloads
ASDTests	Fadi Thabtah	Toddler and adults	Static scoring	4.7	217	√	√	10	5000+
Awesomely Autistic	Android in London	Adults	Static scoring	4.1	54	×	√	50	10000+
Autism Test	Consurgo	Adults	Static scoring	3.3	132	×	√	20	10000+
ANDDS	Apple Inc.	Infants	Static scoring			×	×	Unknown	
ASDetect	Sadka et al.	11 to 30 months	Static scoring	4.3	39	√	√	14	5000+
Autism AI	Seyed Reza Shahamiri and Fadi Thabtah	Toddler and adults	Machine learning	4.2	178	×	√	14	1000+

on that score, the screening outcome is decided. Likewise, ASDTests and Autism AI are the only apps that cover all age groups (infants, children, teenagers, adults), whereas the remaining apps focus on specific age categories.

Awesomely Autistic and Autism Test have high numbers of downloads at the time of this writing since they were published in 2016 and 2015, respectively, but ASDTests has the largest number of reviews besides achieving the highest rating among the considered screening apps. In comparison, Autism AI has achieved a high rating and number of reviews considering it was released in August 2018. In terms of platforms, ASDetect and ASDTests are available for both Android and iOS devices. Most of the apps are only available for English speakers except ASDTests, which is available in ten different languages including French, Turkish, Russian, Spanish, Urdu, Swahili, Arabic, Portuguese, Mandarin, and English, making it the most accessible screening app. In terms of scientific validation, ASDTests is the only application that has been scientifically reviewed and published in an article related to health informatics [26]; hence it is the only application that is academically verified among all (in addition to Autism AI). Nevertheless, instead of Autism AI, all of the aforementioned apps still adopt static rules for the screening process and classic scoring functions for prediction and hence, they can be criticized as being subjective.

The Proposed Autism Detection System: Autism AI

This section explains the proposed Autism AI System and its components. It also details the CNN algorithm, the user experience, and the user interface of the system.

System Architecture

Figure 1 shows the architecture of Autism AI System. It is composed of a mobile app, an Intelligent Autistic Traits Detection web service that enables communications between Autism AI app and the CNN, a database to store the subject's

responses and test results, and the CNN screening algorithm that detects autistic traits, all of which were implemented by the authors and available publicly since August 2018 [29].

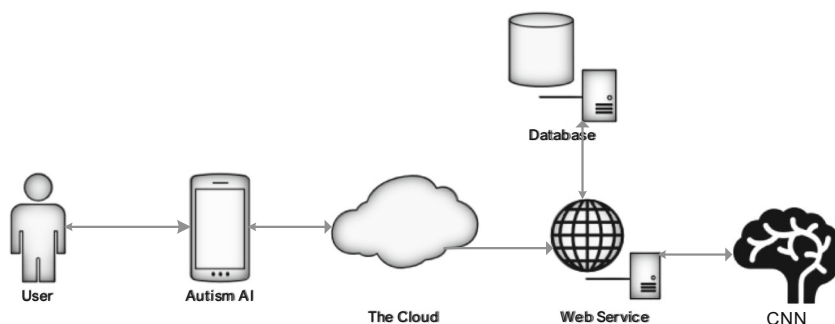
Autism AI app requires to communicate with the web service that interfaces and implements the CNN. The app's responsibility is mainly to provide a professionally designed user interface that is easy to be accessed by caregivers and family members and to provide instant results regarding autistic traits. Moreover, the app captures and verifies relevant user data (behavioral traits and demographic features) and feeds them to the CNN via the web service. Once the user undergoes a screening test and the test result becomes available, Autism AI app also generates a report in which the user can provide to health professionals. The next section provides more information about the user interface.

Autism AI Interface and User Interactions

Figure 2 depicts the activity diagram in which the flow of Autism AI System activities and actors are shown, and Fig. 3 depicts the user interface of Autism AI app. The main users of the system are parents, caregivers, clinicians, medical staff, teachers, and even individuals (adults, adolescents) who have average intelligence quotient (IQ), among others. Autism AI behavioral questions were adopted from Q-CHAT-10 and AQ-10 screening method versions explained in the "Literature Review" section.

Upon launching Autism AI, the first screen is shown that provides some information about the app (Fig. 3a). From here, users can start the Autism Test by selecting whether the test is taken for a toddler less than 36-month-old or an older individual. Since the behavioral questions adopted from Q-CHAT-10 for toddlers are different from adults, adolescents, and children groups, the app requires this information to automatically load the correct behavioral questions and answer options. Once the test is started, the first screen asks questions about test subject's gender, ethnicity, and age (Fig. 3b). If the user is not a toddler, the subject's age is used to put the user in either child, adolescent, or adult age categories automatically, and the app loads the appropriate question set as indicated by AQ-

Fig. 1 Autism AI System architecture



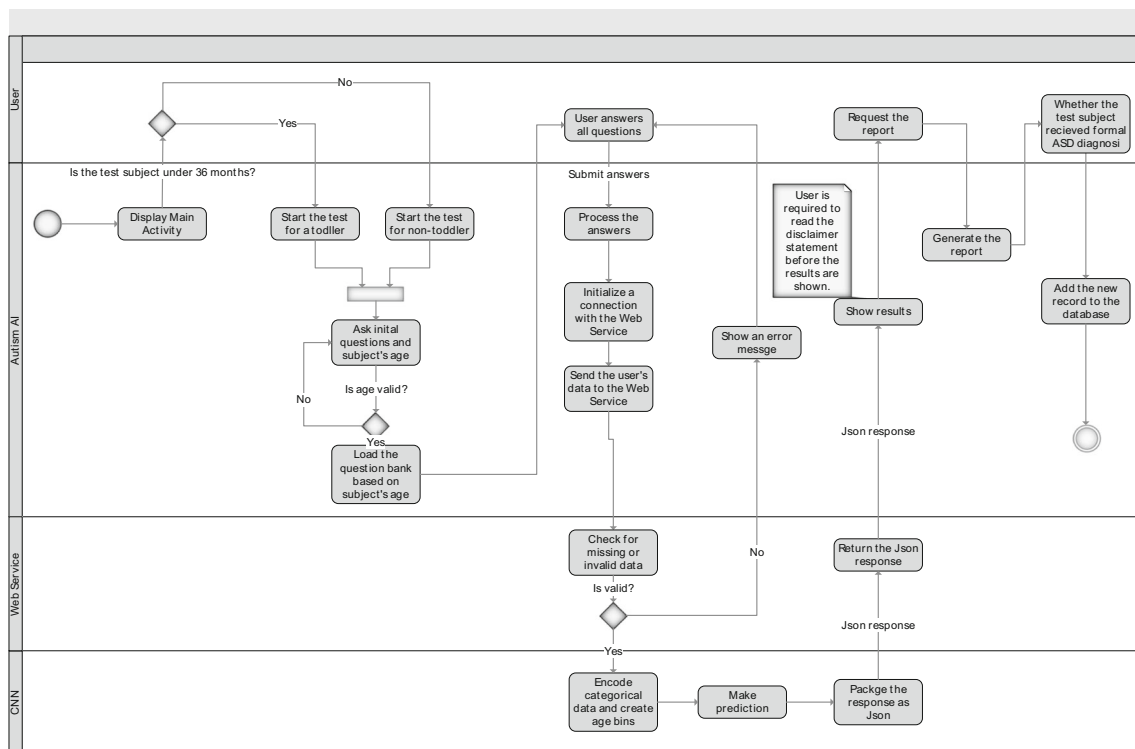


Fig. 2 The proposed system activity diagram

10. An age verification is performed to ensure a valid age is given to the system ($18 \leq \text{age} < 36$ months for toddlers and $3 \leq \text{age} \leq 80$ years for other test subjects).

Once the demographic questions are answered, the app commences asking the ten behavioral questions based on the given age. Each question is displayed in a separate screen for easy navigation, and users can answer each question by selecting one of “definitely agree,” “slightly agree,” “slightly disagree,” and “definitely disagree” options. For toddler test subjects, the answer options are different since Q-CHAT-10 method was employed as the basis of such questions. User responses are then converted into binary representations (0 or 1) by considering the first two options in the possible answers as “1” and the rest as “0” based on the recommendations of AQ-10 and Q-CHAT-10 methods as well as recent research related to autism pre-diagnosis [24, 26]. Users need to answer all the questions and press the submit button (Fig. 3c)—they can also navigate between the questions and change their responses or restart the test entirely.

Once all questions are answered, Autism AI app initiates a connection to the web service by creating an asynchronous task and sends the user’s data to the cloud (Fig. 3d). The web service performs data validation first before passing the data to the CNN algorithm (the CNN is explained in the next section) and informs the app if the data is invalid. Otherwise, the validated data is passed to the CNN to predict the ASD likelihood given the user data and returns the CNN prediction to the web service. It is important to note that the decision of

whether an individual exhibits autistic trait is performed solely by the CNN and not using conventional scoring functions as in AQ and Q-CHAT-10 screening methods. Once the classification decision is made by the detection algorithm, the web service will map the result into a user-readable text and pass it to the app in which the app shows it to the user (Fig. 3e). Nonetheless, before the result is shown, the following disclaimer is presented to the user in which he/she must agree in order to view the result:

“The result provided here is generated by Artificial Intelligence screening tool based on behavioral tests that can pinpoint Autism Spectrum Disorder (ASD) traits and is not diagnosis. If you are concerned that the respondent has ASD, discuss your concerns with a health professional.”

Likewise, participants are required to consent for data use prior to completing the screening; that anonymous user data can only be used for research purposes and are stored in a secured location.

Finally, the test result becomes visible after the user agrees with the disclaimer. From here, the user can either restart the test or download a PDF report (Fig. 3f) for their perusal. There is one last question that the user is required to respond to, and it is whether the test subject has received a formal clinical ASD diagnosis. This question is asked to identify false positives/negatives and improve the CNN predictions in the future.

The app automatically sends anonymous test subject’s data, the CNN prediction, and the formal diagnosis status back to the secured web service to be stored in the database. These

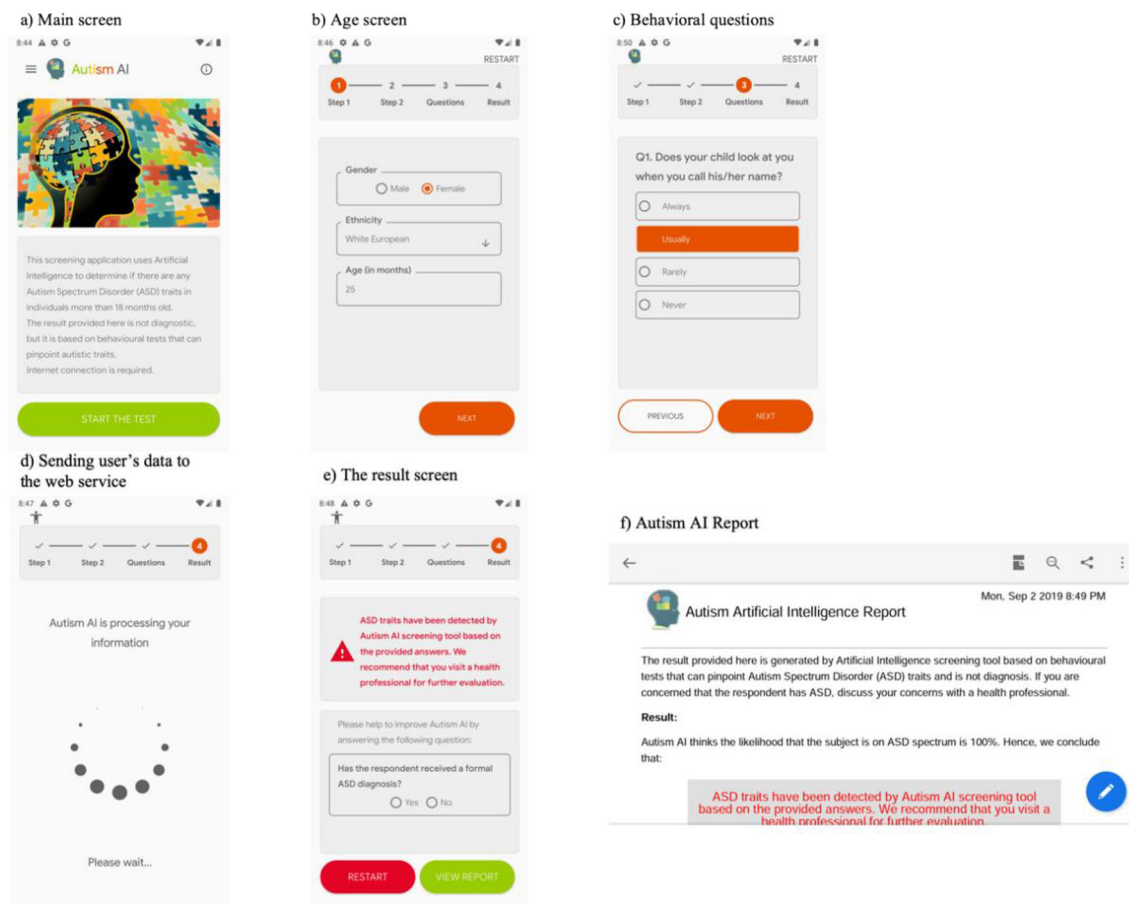


Fig. 3 Autism AI user interface

data become part of the training dataset for future tuning of the prediction algorithm. It is pertinent to highlight that no user identification information is stored in this system and all user information is completely anonymized.

The CNN

The input to the CNN prediction algorithm is the user's data presented as a 1D tensor with 14 coefficients that include user responses to the ten behavioral questions, participant age, gender, whether the user has jaundice, and any ASD history in the family. We merged the datasets from different age publications and compiled one dataset representing the entire data samples collected from different age categories. Primarily, the initial data used to train the CNN was obtained from [24]. The authors of [24] made the data public and obtained an ethical approval from the University of Huddersfield. Since the participant's age is one of the independent variables considered during the CNN training, the CNN learns to properly correlate age, among the other variables, to ASD class when it looks for ASD traits.

This data is then pre-processed by applying one-hot encoding, removing dummy variables, and presenting user

age in bins of size three; the data pre-processing procedure increases the input tensor dimension from 14 to 40 coefficients. These data are then fed to the CNN.

The CNN used here is composed of two convolutional layers with 32 and 64 filters, respectively, followed by a max pooling layer for down-sampling the feature maps after each convolution layer. Since users' data are in form of 1D tensors, we did not apply the standard 2D windows on the feature maps. Particularly, both convolution layers applied a 3×1 window to the feature maps and down-sampling was done by a kernel of size 4×1 . The convolutional layers had no strides (i.e., 1×1), while max pooling strides were 2×2 .

Identifying the dense fully connected hyperparameters was done via a grid search algorithm [31] where 2 to 4 dense layers with 32, 64, and 128 neurons with different activations were trained and verified. Then, the setup with the best performance was selected. As the result of the grid search algorithm, the CNN architecture was selected as shown in Fig. 4; the remaining CNN hyperparameters are provided in Table 2.

We implemented the CNN in Python using Google's TensorFlow library and trained it with the data provided by [28]. After training, the CNN was stored on the Autism AI server so making a prediction only requires loading the pre-trained

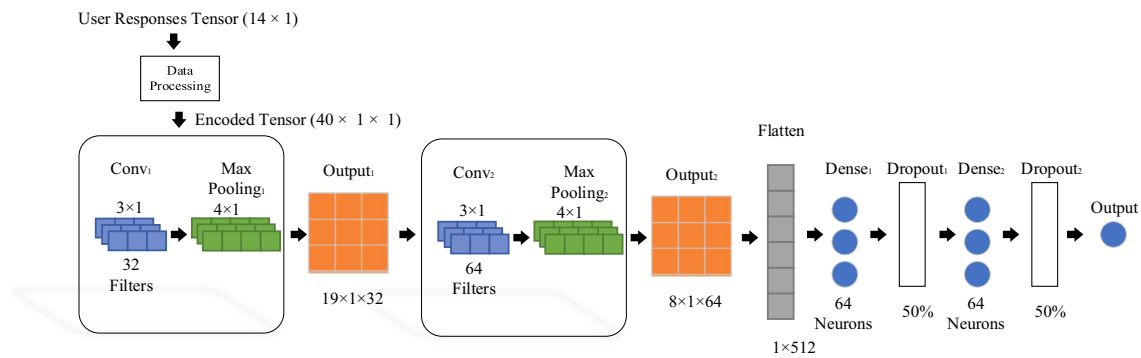


Fig. 4 The CNN architecture

CNN. In addition, the CNN is adaptive, i.e., it fine-tunes and re-trains once a batch of new user data becomes available to include new knowledge captured from new users using the system.

The output of the CNN is the ASD likelihood associated with the user profile in which the CNN returns this prediction result to the user via the web service. The Intelligent Autistic Traits Detection web service is available to the public from [32], and its operations are accessible via the following free application programming interfaces (APIs):

1. */train/*: this function enables the web service admin to retrain the CNN once new user data samples are available and to enable the CNN to take into account the false positives/negatives according to user reports on whether they received formal ASD diagnosis.
2. */predict/*: this function enables users to perform a prediction. It requires the users to supply the following data items in the given order: replies to the ten behavioral questions separated by “,” as zeros and ones, age, gender, ethnicity, jaundice, and family ASD history. All data items are required.
3. */insert_new_row/*: this function enables authorized users to insert a new test sample to the database. In addition to the data required for *predict*, the CNN prediction and formal ASD diagnosis status must be sent as well. This function is automatically called by the app.

Autism AI app is available for Android users from Google Play Store since August 2018 from [30]. The next section explains how the CNN was evaluated.

Results Analysis and Evaluation

In verifying the CNN prediction algorithm, we selected a repeated random subsampling cross-validation procedure with tenfolds. Furthermore, other machine learning algorithms were considered and evaluated using the same data for comparison purposes. In each cross-validation fold, the dataset was randomly shuffled; then 75% of the data were used for CNN training and the rest to test its performance. The dataset was reshuffled per each fold.

We measured the following metrics to evaluate the performance of the CNN and compare with the other prediction algorithms considered here:

1. Accuracy (%): accuracy was measured as the ratio of correct classifications to the number of total tests. If the CNN likelihood prediction is more than 50%, the system returns true as the response to the user that means ASD traits have been detected in the subject and false otherwise:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{n} \quad (1)$$

where n is the number of total tests per fold.

2. Sensitivity (%): sensitivity (and specificity) is a binary classification metric commonly used to verify medical tests and screening studies. It provides the proportion of tests that are correctly classified as true positive. To put it differently, it is the ratio of subjects with ASD correctly identified. It was calculated as:

Table 2 The CNN hyperparameters

Layers	Activation function	Dropout	Other parameters
Both convolutional layers	Relu	NA	Optimizer: Adam
Both dense layers: 64 neurons each	Elu	50%	Learning rate α : 0.001
Output: one neuron	Sigmoid (to perform binary classification)	NA	Loss function: binary cross entropy Mini-batch: 256

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (2)$$

3. Specificity (%): similar to sensitivity, specificity provides the ratio of tests that are correctly classified as true negative, i.e., the proportion of subjects without ASD that were correctly classified as healthy:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (3)$$

The dataset included overall 6075 samples in which 4556 random samples were used during training and 1519 for testing in each fold. The data contained 42% female subjects out of which 31% were identified with ASD. Likewise, 30% of the male participants were identified with ASD. Out of 1045 subjects with jaundice, 29% of them were participants with autistic traits as well. There were 81% of the participants without any autistic family history in which 31% of them were autistic, while only 25% of the subjects with ASD family history were classified to have ASD. Overall, 69% of the participants were labeled with no autistic traits detected and 31% otherwise. To remedy this class imbalance problem, the class weight of the training samples that represented individuals with autistic traits was increased, while smaller weights were given to other samples during the CNN training, as explained in [33]. This approach instructs the CNN prediction algorithm to pay increased attention to samples with positive autistic traits. Table 3 provides the cross-validation results obtained from the CNN.

The CNN delivered average testing accuracy of 97.95% with a mean sensitivity of 95.53% and specificity of 98.63%. The training accuracies were also on a par with the testing accuracies that show the lack of overfitting since the applied dropout regularization technique limited the memory capacity of the CNN and pushing it towards learning the autistic patterns.

Furthermore, the statistical analysis of the CNN's specificity and sensitivity according to their standard deviations σ is presented by Fig. 5. The blue lines depict one σ distance from the mean ($\pm 1\sigma$) in both directions, while the red lines are two σ ($\pm 2\sigma$). The ten red dots in each graph are specificity and sensitivity observations obtained from cross-validation folds given by Table 4. For specificity, 70% of the observations were between $\pm 1\sigma$ and the rest between $\pm 2\sigma$. For sensitivity, 80% fall between $\pm 1\sigma$ and two observations between $\pm 2\sigma$. Thus, it can be concluded that both sensitivity and specificity results follow a normal distribution. There was also no outlier observation. These are other indications of lack of overfitting in the results presented by Table 3.

The machine learning algorithms selected for comparison were Ripple Down Rule learner (Ridor), Bayes Net, and C4.5 Decision Tree. The reason for using these algorithms is that they utilize different learning methods when processing data. For example, C4.5 employs information gain principle in constructing classifiers, Bayes Net uses Bayes' theorem, whereas Ridor uses rule induction approach to form rules.

Table 4 depicts predictive mean accuracy, sensitivity, and specificity results derived by the machine learning algorithms against the autism dataset. Among the three ML algorithms, CNN provided the best results for all evaluation metrics, while C4.5 improvements are marginal to Ridor. However, both C4.5 and Ridor performances were by far better than Bayes Net when around 7% better accuracy and sensitivity and about 10% better specificity were achieved. Compared with the proposed CNN in Autism AI System, the considered machine learning algorithms performed poorly with respect to all evaluation metrics mentioned before. For instance, CNN produced 7.62%, 17.07%, and 9.35% higher predictive accuracy than C4.5, Bayes Net, and Ridor, respectively.

The increase in accuracy rate is attributed to the learning scheme employed by the CNN. Particularly, the ability of CNNs to progressively increase feature maps depth means that they can present the original autistic data in more detailed

Table 3 The CNN evaluation results

Fold	Training loss (cross entropy)	Testing accuracy (%)	Testing sensitivity (%)	Testing specificity (%)
1	0.0260	97.95	98.63	96.53
2	0.0205	98.28	99.52	96.54
3	0.0327	98.29	98.59	97.57
4	0.0209	97.96	98.41	96.89
5	0.0288	98.15	98.39	97.59
6	0.0256	98.42	98.58	98.03
7	0.0232	98.61	98.97	97.73
8	0.0261	98.48	98.83	97.75
9	0.0255	98.55	98.50	98.66
10	0.0223	98.88	99.60	97.17

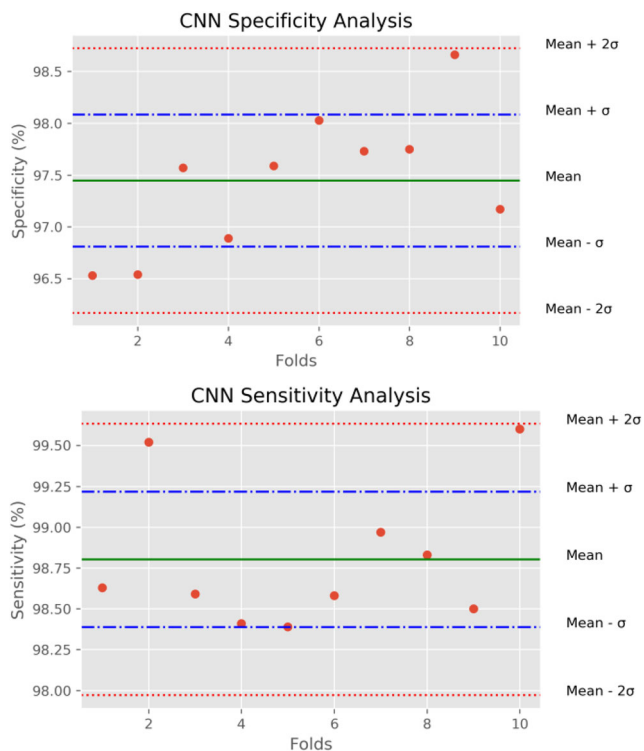


Fig. 5 Sensitivity and specificity analysis

ways and extract more effective features and coefficients. To put it differently, the original 40 coefficients initially fed to the CNN were transformed into 512 coefficients using the two convolutional components of 32 and 64 filters plus pooling functions. Likewise, the CNN’s dense layers were successful in learning the hidden knowledge extracted by the previous convolutional operations.

The sensitivity and specificity rates of the CNN were similarly higher than those derived by the ML algorithms on the autism dataset. In particular, the CNN improved autism classification sensitivity by 6.23%, 15.63%, and 7.93% with respect to the C4.5, Bayes Net, and Ridor algorithms, respectively. All the ML algorithms considered here showed acceptable level of sensitivity rates as they achieved figures higher than 80%. For specificity, they reported good rates as well although Bayes Net reported lower than expected specificity rate due to the high false positives generated by this algorithm

Table 4 Performance comparative study

Algorithm	Testing accuracy (%)	Testing specificity (%)	Testing sensitivity (%)
<i>C4.5</i>	90.33	94.65	90.30
<i>Bayes Net</i>	80.88	83.90	80.90
<i>Ridor</i>	88.60	96.18	88.60
<i>CNN</i>	97.95	98.63	96.53

against the autism dataset; to be specific, Bayes Net misclassified 678 instances with no autistic traits. On the other hand, C4.5 and Ridor derived only 225 and 161 false positives, respectively, when processing the autism dataset. This indeed pinpointed that Bayes Net is the least appropriate classification method for autism detection at least on the dataset we have utilized.

Based on the confusion matrix results (true positives, true negatives, false positives, and false negatives), it was apparent that C4.5 derived lower false negatives than the remaining algorithms except the CNN. The false negatives represent individuals that are wrongly classified to be without autism when there were actually on the spectrum. Based on the results generated, C4.5 wrongly classified 19.44% of the test subjects to be without ASD, whereas Ridor and Bayes Net algorithms incorrectly classified 28.51% and 25.95%, respectively. These misclassifications had negatively impacted the true positive rate for class label “ASD = Yes” where C4.5, Bayes Net, and Ridor derived 80.60%, 74.10%, and 71.50% true positive rates for class “ASD = yes.” However, for class “ASD = No”, C4.5, Bayes Net, and Ridor derived 94.70%, 83.90%, and 96.20% true positive rates, respectively. Taking the average of both class labels, the true positive rates generated by C4.5, Bayes Net, and Ridor against the autism dataset are 90.30%, 80.90%, and 88.60% respectively. On the other hand, the CNN achieved higher true positive rates on both classes than the considered ML algorithms in every fold. In particular, the CNN delivered in average 95.99% true positive for class “ASD = yes” and 99.15% for class “ASD = No” that means 97.57% true positive across both classes. This shows significant improvements offered by the CNN in detecting ASD traits compare with the ML algorithms where up to 24.49% better ASD detection was achieved for subjects with autism.

Thus, CNNs are more capable to provide a precise model of ASD; hence they can be used to detect autistic traits with higher accuracy. The superiority of the CNN is clear in all results derived against the considered dataset with reference to accuracy, sensitivity, and specificity. This is because of (1) the CNNs’ characteristic of applying multiple filters to represent the original features explained before and (2) the CNNs’ capability in learning translation invariant features. Particularly, CNNs learn local features compare with other algorithms that learn global features. This means once the CNN learns patterns associated with ASD in any location of its feature space, it is still capable of detecting that ASD pattern if they occur anywhere else in the feature space, while other algorithms are incapable to do so. Additionally, behavioral imaging would enable to decrease the subjectivity of questionnaire-based data and thus increase reproducibility of features taken into account by the CNN [33]. Autism AI is the first ASD screening system that employs such technique and enables non-specially trained professionals to leverage deep learning technologies.

It is pertinent to note that, with respect to the validation question at the end of the questionnaire (whether the test subject has received a formal clinical ASD diagnosis), the user replies are highly consensual but not perfect and slightly heterogeneous and the question does not indicate how the diagnosis was done (e.g., professional or gold standard evaluation ADI-ADOS). This is a limitation especially in the case of a spectrum rather than a dichotomous clear clinician classification. We will refine this question in the next iterations of system refinements to overcome this limitation.

Conclusions

Emerging technologies, such as deep learning, provide end-users and decision-makers with the powerful capabilities in data analytics and visualization that indeed can improve the quality and effectiveness of decision-making. This research proposed a new autism detection system called Autism AI that employs CNNs for autism screening. Autism AI System is not just a conventional mobile application for screening since prediction is primarily based on learning from cases and controls rather than scoring functions based on specific rules. In coming up with a prediction decision, Autism AI processes data provided by individuals or their families or medical professionals using a CNN classification algorithm to quickly detect the possibility of autistic traits. The interaction with Autism AI System is achieved via a well-designed mobile application with easy to navigate graphical user interface linked to multiple Web APIs and a database on a secured cloud. The Autism AI System can be accessible from both Google and Apple Play Stores, and its APIs are available for developers globally to bring its functionalities to other platforms and applications.

Experimental results using over 6000 instances and several machine learning algorithms showed that the CNN detection algorithm was superior when compared with decision trees, rule induction, and Bayes Net algorithms. The comparison was conducted based on different evaluation metrics such as accuracy, sensitivity, specificity, and true positives rates. The results derived showed that the proposed CNN algorithm within the Autism AI derived higher predictive accuracy than the considered machine learning algorithms.

Currently, Autism AI has 4.2 out of 5 review rates on Google Play Store with 178 user reviews, while more than 3000 ASD screening tests were conducted using the proposed system since its public launch in August 2018 and the time of this writing. The majority of the reviews are positive commenting on app's ease of use, usability, and esthetic and confirming if the system delivered correct diagnosis. Nevertheless, there have been few negative reviews mostly concerning the use of artificial intelligence for ASD screening and accusing the system does not actually use AI. There was one user indicating that the system did not recognize he was

on the spectrum. Since there was no way for us to identify the information supplied by this user to the app, we requested him/her to send us the report so that we could investigate further, which he did not reply. These reviews are publicly available on app's Google Play Store page [29].

One of the limitations of this study is the exclusion of complex features such as videos and images related to cases and controls. In the near future, the proposed AI screening system can be expanded to possibly explore advanced deep learning schemes that can detect new unconventional features of autism from complex features. Furthermore, future studies can investigate cluster analysis to identify endophenotypes, assess the role of development to help the diagnosis (since some features are more important for children or adults), and refine the prognosis and the therapeutic strategy.

Compliance with Ethical Standards

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest Seyed Reza Shahamiri declares that he has no conflict of interest. Fadi Thabtah declares that he has no conflict of interest.

References

1. S. A. Schoen *et al.*, "A systematic review of Ayres sensory integration intervention for children with autism," *Autism Research*. 2019.
2. Towle PO, Patrick PA. Autism spectrum disorder screening instruments for very young children: a systematic review: *Autism Res. Treat*; 2016.
3. F. Thabtah and Fadi, "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment," in *Proceedings of the 1st International Conference on Medical and Health Informatics 2017 - ICMHI '17*, 2017.
4. Wall DP, Dally R, Luyster R, Jung JY, DeLuca TF. Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism. In: Use of artificial intelligence to shorten the behavioral diagnosis of autism: *PLoS One*; 2012.
5. M. Duda, J. A. Kosmicki, and D. P. Wall, "Testing the accuracy of an observation-based classifier for rapid detection of autism risk," *Translational psychiatry*. 2015.
6. M. Marlow, C. Servili, and M. Tomlinson, "A review of screening tools for the identification of autism spectrum disorders and developmental delay in infants and young children: recommendations for use in low- and middle-income countries," *Autism Research*. 2019.
7. Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioral distinction of autism and ADHD. In: Use of machine learning for behavioral distinction of autism and ADHD. *Psychiatry: Transl*; 2016.
8. 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*. Aug. 2016;57(8): 927–37.
9. Thabtah F. Machine learning in autistic spectrum disorder behavioral research: a review and ways forward. *Care: Informatics Heal. Soc*; 2018.

10. Little SG, Akin-Little A, Harris GM. Autism spectrum disorder: screening and diagnosis. In: *Behavioral interventions in schools: Evidence-based positive strategies*. 2nd ed; 2019.
11. S. R. Shahamiri and S. S. B. Salim, “A multi-views multi-learners approach towards dysarthric speech recognition using multi-nets artificial neural networks,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 1053–1063, 2014, A Multi-Views Multi-Learners Approach Towards Dysarthric Speech Recognition Using Multi-Nets Artificial Neural Networks.
12. Sremath S, et al. Speaker identification features extraction methods : a systematic review. *Expert Syst. Appl.* 2017;90:250–71.
13. Baron-Cohen S, Allen J, Gillberg C. Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *Br. J. Psychiatry.* 1992;161:839–43.
14. Robins DL, Fein D, Barton ML, Green JA. The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders: *J. Autism Dev. Disord*; 2001.
15. C. Allison, B. Auyeung, and S. Baron-Cohen, “Toward brief ‘Red Flags’ for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls,” *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 51, no. 2, pp. 202–212.e7, Feb. 2012.
16. Wong V, et al. A Modified Screening Tool for Autism (Checklist for Autism in Toddlers [CHAT-23]) for Chinese Children. In: A modified screening tool for autism (Checklist for Autism in Toddlers [CHAT-23]) for Chinese children: *Pediatrics*; 2004.
17. 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.* Feb. 2001;31(1):5–17.
18. Wheelwright S, et al. Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). In: Predicting autism spectrum quotient (AQ) from the systemizing quotient-revised (SQ-R) and empathy quotient (EQ): *Brain Res*; 2006.
19. Auyeung B, Baron-Cohen S, Wheelwright S, Allison C. The autism spectrum quotient: children’s version (AQ-Child): *J. Autism Dev. Disord*; 2008.
20. K.-C. Chu, H.-J. Huang, and Y.-S. Huang, “Machine learning approach for distinction of ADHD and OSA,” in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016, pp. 1044–1049.
21. T. Wolfers, J. K. Buitelaar, C. F. Beckmann, B. Franke, and A. F. Marquand, “From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics,” *Neuroscience and Biobehavioral Reviews*. 2015.
22. J. L. Marcano, M. A. Bell, and A. A. (Louis. Beex, “Classification of ADHD and non-ADHD subjects using a universal background model,” *Biomed. Signal Process. Control*, 2018, Classification of ADHD and non-ADHD subjects using a universal background model.
23. M. J. Maenner, M. Yeargin-Allsopp, K. Van Naarden Braun, D. L. Christensen, and L. A. Schieve, “Development of a machine learning algorithm for the surveillance of autism spectrum disorder,” *PLoS One*, vol. 11, no. 12, p. e0168224, Dec. 2016.
24. Thabtah F, Kamalov F, Rajab K. A new computational intelligence approach to detect autistic features for autism screening. In: A new computational intelligence approach to detect autistic features for autism screening: *Int. J. Med. Inform*; 2018.
25. F. Thabtah, “An accessible and efficient autism screening method for behavioural data and predictive analyses,” *Health Informatics J.*, p. 146045821879663, Sep. 2018.
26. Ltd HM. Awesomely Autistic Test: Google Play Store; 2016.
27. “Autism Test.” Google Play Store, 2015.
28. N. Sadka, W. Nadachowski, C. Dissanayake, and J. Barbaro, “Early childhood autism surveillance and assessment tool | ASDetect.” 2016.
29. S. R. Shahamiri and F. Thabtah, “Autism AI,” 2018. [Online]. Available: [https://play.google.com/store/apps/details?id = com.rezanet.intelligentasdscreener](https://play.google.com/store/apps/details?id=com.rezanet.intelligentasdscreener).
30. Shahamiri SR, Wan-Kadir WMN, Ibrahim S, Hashim SZM. Artificial neural networks as multi-networks automated test oracle. *Autom. Softw. Eng.* 2012;19(3):303–34.
31. S. R. Shahamiri and F. Thabtah, “Intelligent Autistic Traits Detection Web Service,” 2018. [Online]. Available: <http://rshahamiri.pythonanywhere.com/>.
32. Dong Q, Gong S, Zhu X. Imbalanced deep learning by minority class incremental rectification. *IEEE Trans. Pattern Anal. Mach. Intell.* 2018:1–1.
33. Egger HL, et al. Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study. *npj Digit. Med.* Dec. 2018;1(1):1–10.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.