



Diagnosing Autism Spectrum Disorder Without Expertise: A Pilot Study of 5- to 17-Year-Old Individuals Using Gazefinder

Kenji J. Tsuchiya^{1,2*}, Shuji Hakoshima³, Takeshi Hara^{4,5}, Masaru Ninomiya³, Manabu Saito^{6,7}, Toru Fujioka^{2,8,9}, Hirotsuka Kosaka^{2,9,10}, Yoshiyuki Hirano^{2,11}, Muneaki Matsuo¹², Mitsuru Kikuchi^{2,13,14}, Yoshihiro Maegaki¹⁵, Taeko Harada^{1,2}, Tomoko Nishimura^{1,2} and Taiichi Katayama²

¹ Research Center for Child Mental Development, Hamamatsu University School of Medicine, Hamamatsu, Japan, ² Department of Child Development, United Graduate School of Child Development, Osaka University, Kanazawa University, Hamamatsu University School of Medicine, Chiba University, and University of Fukui, Suita, Japan, ³ Healthcare Business Division, Development Center, JVCKENWOOD Corporation, Yokohama, Japan, ⁴ Center for Healthcare Information Technology, Tokai National Higher Education and Research System, Gifu, Japan, ⁵ Faculty of Engineering, Gifu University, Gifu, Japan, ⁶ Department of Neuropsychiatry, Graduate School of Medicine, Hirosaki University, Hirosaki, Japan, ⁷ Research Center for Child Mental Development, Graduate School of Medicine, Hirosaki University, Hirosaki, Japan, ⁸ Department of Science of Human Development, Faculty of Education, Humanities and Social Sciences, University of Fukui, Fukui, Japan, ⁹ Research Center for Child Mental Development, University of Fukui, Fukui, Japan, ¹⁰ Department of Neuropsychiatry, Faculty of Medical Sciences, University of Fukui, Fukui, Japan, ¹¹ Research Center for Child Mental Development, Chiba University, Chiba, Japan, ¹² Department of Pediatrics, Faculty of Medicine, Saga University, Saga, Japan, ¹³ Department of Psychiatry and Neurobiology, Graduate School of Medical Science, Kanazawa University, Kanazawa, Japan, ¹⁴ Research Center for Child Mental Development, Kanazawa University, Kanazawa, Japan, ¹⁵ Tottori University Hospital, Yonago, Japan

OPEN ACCESS

Edited by:

Josephine Barbaro,
La Trobe University, Australia

Reviewed by:

Francesca Felicia Operto,
University of Salerno, Italy
D. Mishra,
University of Delhi, India
Maurizio Elia,
Oasi Research Institute (IRCCS), Italy

*Correspondence:

Kenji J. Tsuchiya
tsuchiya@hama-med.ac.jp

Specialty section:

This article was submitted to
Pediatric Neurology,
a section of the journal
Frontiers in Neurology

Received: 05 September 2020

Accepted: 30 December 2020

Published: 28 January 2021

Citation:

Tsuchiya KJ, Hakoshima S, Hara T, Ninomiya M, Saito M, Fujioka T, Kosaka H, Hirano Y, Matsuo M, Kikuchi M, Maegaki Y, Harada T, Nishimura T and Katayama T (2021) Diagnosing Autism Spectrum Disorder Without Expertise: A Pilot Study of 5- to 17-Year-Old Individuals Using Gazefinder. *Front. Neurol.* 11:603085. doi: 10.3389/fneur.2020.603085

Atypical eye gaze is an established clinical sign in the diagnosis of autism spectrum disorder (ASD). We propose a computerized diagnostic algorithm for ASD, applicable to children and adolescents aged between 5 and 17 years using Gazefinder, a system where a set of devices to capture eye gaze patterns and stimulus movie clips are equipped in a personal computer with a monitor. We enrolled 222 individuals aged 5–17 years at seven research facilities in Japan. Among them, we extracted 39 individuals with ASD without any comorbid neurodevelopmental abnormalities (ASD group), 102 typically developing individuals (TD group), and an independent sample of 24 individuals (the second control group). All participants underwent psychoneurological and diagnostic assessments, including the Autism Diagnostic Observation Schedule, second edition, and an examination with Gazefinder (2 min). To enhance the predictive validity, a best-fit diagnostic algorithm of computationally selected attributes originally extracted from Gazefinder was proposed. The inputs were classified automatically into either ASD or TD groups, based on the attribute values. We cross-validated the algorithm using the leave-one-out method in the ASD and TD groups and tested the predictability in the second control group. The best-fit algorithm showed an area under curve (AUC) of 0.84, and the sensitivity, specificity, and accuracy were 74, 80, and 78%, respectively. The AUC for the cross-validation was 0.74 and that for validation in the second control group was 0.91. We confirmed that the diagnostic performance of the best-fit algorithm is comparable to the diagnostic assessment tools for ASD.

Keywords: autism spectrum disorder, school-age children, adolescent, Gazefinder, machine learning, Japan

INTRODUCTION

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder characterized by atypicality in social communication, and restricted and repetitive behaviors. A recent epidemiological study from Japan reported that the prevalence of ASD is higher than 3% in the general population at the age of 5 years (1). ASD affects the quality of life across the lifespan of the affected individual (2). Various early intervention practices have been developed, some of which are effective and promising (3). Timely intervention is key to better outcomes (4), for which accurate diagnostic assessment is prerequisite.

Regardless of the clinical significance of diagnosis, professionals face challenges inherent to diagnostic assessments of ASD. The first challenge lies in the nature of the diagnosis. As there are no well-established biophysiological diagnostic tests, diagnosis is made solely based on behavioral assessment of children. However, the development of such signs in children is not stable along the time course of diagnosis and can change as they grow. The complexity of social engagement in children generally increases from month to month, especially during a younger age. Furthermore, complex social interactions are affected by the children's cooperativeness, which is further influenced by physical conditions such as hunger and fatigue, as well as their moods and tempers (5–7). Furthermore, no single behavioral sign or trait sufficiently points toward the diagnosis. Therefore, a single diagnostic assessment is usually insufficient to confirm the diagnosis of ASD. To overcome this challenge, standardized tools for diagnosing ASD have been developed, including the Autism Diagnostic Observation Schedule, second edition (ADOS-2) (8) and the Autism Diagnostic Interview-Revised (9), which are widely accepted in research and clinical settings because of their high reliability and validity. However, the use of these tools leads to the second challenge because an interview with these tools followed by the *post-hoc* assessment takes considerable time. Moreover, interviewers need to have the required expertise and undergo training sessions beforehand. Unfortunately, the costs associated with the use of the above methods have restricted the clinical availability of the tools. A recent Australian study reported that only a small proportion of children were assessed using these tools when parents raised concerns over the possibility of their children having ASD, and a “wait-and-see” approach was advised instead (10). This has likely happened in Japan as well, where only 32% of children confirmed to have a diagnosis of ASD at 5 years had had a history of clinical diagnosis of ASD until the fifth birthday (1). Many children with ASD are left undiagnosed and are not provided appropriate interventions even at school age.

Owing to these two challenges that clinicians face in the diagnostic assessment of ASD, models that balance the quality and accuracy of assessment with timeliness and ease are desired (11). To meet this demand, biological/physiological biomarkers have been tested in several studies.

Among them, atypical eye gaze patterns in children with ASD have been tested to determine whether they can serve as candidate markers. Recent advances in eye gaze studies rely heavily not only on advances in technology, but also on the

fact that eye gaze patterns reflect both biological and behavioral aspects of ASD (11). Eye gaze patterns are under genetic control (12), and a lack of eye gaze onto human faces measured by eye-tracking devices can be a reflection of lack of eye contact with the examiner, a well-known behavioral diagnostic marker of ASD (13). A recent systematic review pointed out that the effect size resulting from the atypicality of eye gaze patterns in individuals with ASD can have standard deviations (SDs) as large as 0.5 (14, 15). The most consistent finding in these review articles is that individuals with ASD spend looking at the non-social stimuli for longer durations than at the social stimuli, human faces in particular (*social paradigm*); the contrast was independent of age, sex, intelligence quotient (IQ), and other conditions. This is consistent with the social processing theory of ASD (14). This atypical eye gaze pattern was more specifically tested in a preferential viewing paradigm, in which individuals, particularly young children with ASD stare preferentially at geometric figures than at human figures (*preferential paradigm*) (16, 17).

Considering the relative ease of measurement and the biophysiological significance of eye gaze, the atypical eye gaze patterns measured with automated eye-tracking devices can serve as diagnostic markers. In preliminary attempts involving young children, the diagnostic validity and test-retest reliability of eye gaze measurements have been supported (16, 17). However, to the best of our knowledge, no study has tested individuals with ASD in a wide age range. The lack of knowledge is particularly evident among school-aged children and adolescents. Furthermore, previous studies have used eye-tracking devices not specifically developed for individuals with ASD. Since some individuals, especially of young ages, cannot cooperate in keeping their eyes on the monitor, the quickest calibration without any intentional cooperation of the child and minimal duration of stimulus movies should be implemented to validate these attempts in clinical settings. To address this, we have attempted to establish specific eye gaze patterns as a biophysiological markers predicting the diagnosis of ASD, using an eye-tracking system called “Gazefinder” in a broad age range of study participants (18, 19). Novel devices and software were designed, including a calibration movie (5–7 s) and stimulus movies. In the stimulus movies, two paradigms were adopted to test the diagnostic predictability of ASD: the social paradigm and the preferential paradigm. The application of Gazefinder to children does not require any expertise, and it takes <2 min to obtain an output (18, 19). Thus, this system is anticipated to fulfill current demands in ASD diagnosis.

In this study, we propose a computerized diagnostic algorithm for ASD using Gazefinder, implemented with social and preferential paradigms in individuals aged 5–17 years. To realize this, we conducted a multisite study to create a computerized best-fit diagnostic algorithm with satisfactory sensitivity and specificity, and validated it in two ways.

METHODS

Participants

Two hundred and twenty-two individuals aged between 5 and 17 years were enrolled by physicians at seven research

sites and affiliated clinics at Hamamatsu University School of Medicine, Hirosaki University, University of Fukui, Chiba University, Saga University, Kanazawa University, and Tottori University during a 6-month period beginning on 25 February 2018. The seven university clinics are located in small or middle-sized cities and metropolitan areas throughout Japan. All clinics play pivotal roles in providing services for children and adolescents with developmental disabilities in the context of child and adolescent psychiatry and/or pediatric neurology. The reasons for enrolling the participating individuals were as follows: (1) They were previously suspected by psychologists, speech therapists, physicians, or school teachers as having ASD, including “autism” and “Asperger disorder,” or (2) they self-nominated to participate in response to the web-based advertisement and have never been suspected to have developmental disorders such as ASD, attention-deficit hyperactivity disorder (ADHD), and learning disabilities. All the participating individuals and their parents were of Japanese ethnicity.

All the legal guardians (i.e., parents in this study) of the participants provided written informed consent, and the participating individuals provided informed assent orally. The study protocol was approved by the ethics committees of the seven research sites and conformed to the tenets of the Declaration of Helsinki.

Measurement

Clinical Evaluation, Screening, and Diagnosis

The initial clinical evaluation by a board-certified psychiatrist or pediatrician included face-to-face behavioral assessment and collection of the developmental history, physical morbidity, and history of medication. This clinical evaluation was followed by screening for ASD using the Pervasive Developmental Disorders Autism Society Japan Rating Scale [PARS, a questionnaire for parents, 57 items (20)], the Strength and Difficulty Questionnaire [SDQ, a questionnaire for parents, 25 items (21)], and the Social Responsiveness Scale in Japanese, second version [SRS-2, a questionnaire for parents, 65 items (22)]. ADHD was screened using the ADHD Rating Scale [ADHD-RS, a questionnaire for parents, 18 items (23)]. General cognitive ability was assessed as indexed by the IQ with the Wechsler Intelligence Scale for Children-fourth edition (WISC-IV) for 215 (97%) individuals, or with the Tanaka-Binet test (Japanese version of the Stanford-Binet Test) for four individuals (2%), or as indexed by developmental quotient (DQ) with the Kyoto Scale of Psychological Development by trained clinical psychologists for three individuals (1%), depending on the participants' mental age. An IQ or DQ lower than 70 was defined as general cognitive delay. An IQ of lower than 70 in WISC-IV is an indication of 2 SD below the population average. The comparability of the IQs derived from the Tanaka-Binet test was tested with the IQ derived from WISC-III, the prior version of WISC-IV (24), and the comparability of the DQs derived from the Kyoto Scale of Psychological Development was tested with the Tanaka-Binet IQ (25).

After the screening tests and assessment of general cognitive abilities, individuals exhibiting positive results to any one of the

three screening tests for ASD (PARS, SDQ, SRS-2) were assessed to have a diagnosis of ASD using the ADOS-2 (a semi-structured, play-based observational assessment tool involving interaction with the child and observation of the activities proposed to the child, 11–16 diagnostic algorithm items depending on module) (8), or the Autism Diagnostic Interview-Revised-Japanese Version (ADI-R-JV, a semi-structured interview for parents, 93 items) (9) conducted by trained clinical researchers. In addition, all the participants were assessed to have a diagnosis of ADHD using Conners 3 Japanese version (Conners 3, a questionnaire for parents, 108 items) (26). We did not use both ADOS-2 and ADI-R-JV in our study because of the limited availability of examiners for these tools who had established research reliability with the developers of the instruments. Among 102 individuals with a positive screening result, 81 individuals were assessed only with ADOS-2, nine individuals were assessed only with ADI-R-JV, and 12 individuals were assessed with both ADOS-2 and ADI-R-JV. The remaining 120 individuals were not assessed with ADOS-2 or ADI-R-JV because of negative results in the screening tests.

Apparatus

Gazefinder, a system in which a set of devices to capture eye gaze patterns and stimulating movie clips are equipped in a personal computer (PC) with a 19-inch monitor (1280 × 1024 pixels), manufactured by JVC Kenwood Co. Ltd. (Yokohama, Japan), was chosen to measure eye gaze patterns. The technical features of this system have been described in previous studies (18, 19, 27, 28). In brief, corneal reflection techniques enable the device to calculate eye gaze positions on the PC monitor as (X, Y) coordinates in pixel units, at a frequency of 50 Hz (i.e., 3,000 eye gaze positions detected per minute). Using the device, the (X, Y) coordinates provide information regarding where a participating child exactly looked at in the movie clip every 1/50 s when the movie clips were shown to the child on the monitor. The participants were asked to sit in front of the monitor to retain the distance between the face and the monitor at approximately 60 cm. Before the diagnostic measurement of the eye gaze patterns, the calibration of the eye gaze position started automatically and took 5–7 s to complete. The calibration was judged as successful if the child looked at no less than three of the five regions of the calibration movie clip. Otherwise, the calibration movie clip restarted. This movie clip contains a black background with a circular region covered with animated animals, moving from the center to the four corners of the monitor consecutively. After the calibration, 12 short movie clips were automatically presented as experimental stimuli in a fixed order. Between the movie clips, we inserted short attention-grabbing movie clips (2.0 s) to set the eye gaze position at the center of the monitor, prior to the stimuli presentation appearing next. This insertion also canceled out the post-images of the movie clip shown previously. The total time of the movie sequence was 95 s, including time for non-stimulus movie clips.

Movie Clips as Stimuli

Figure 1A presents the 12 movie clips used as stimuli. The social paradigm represented in three movie clips were as follows: a

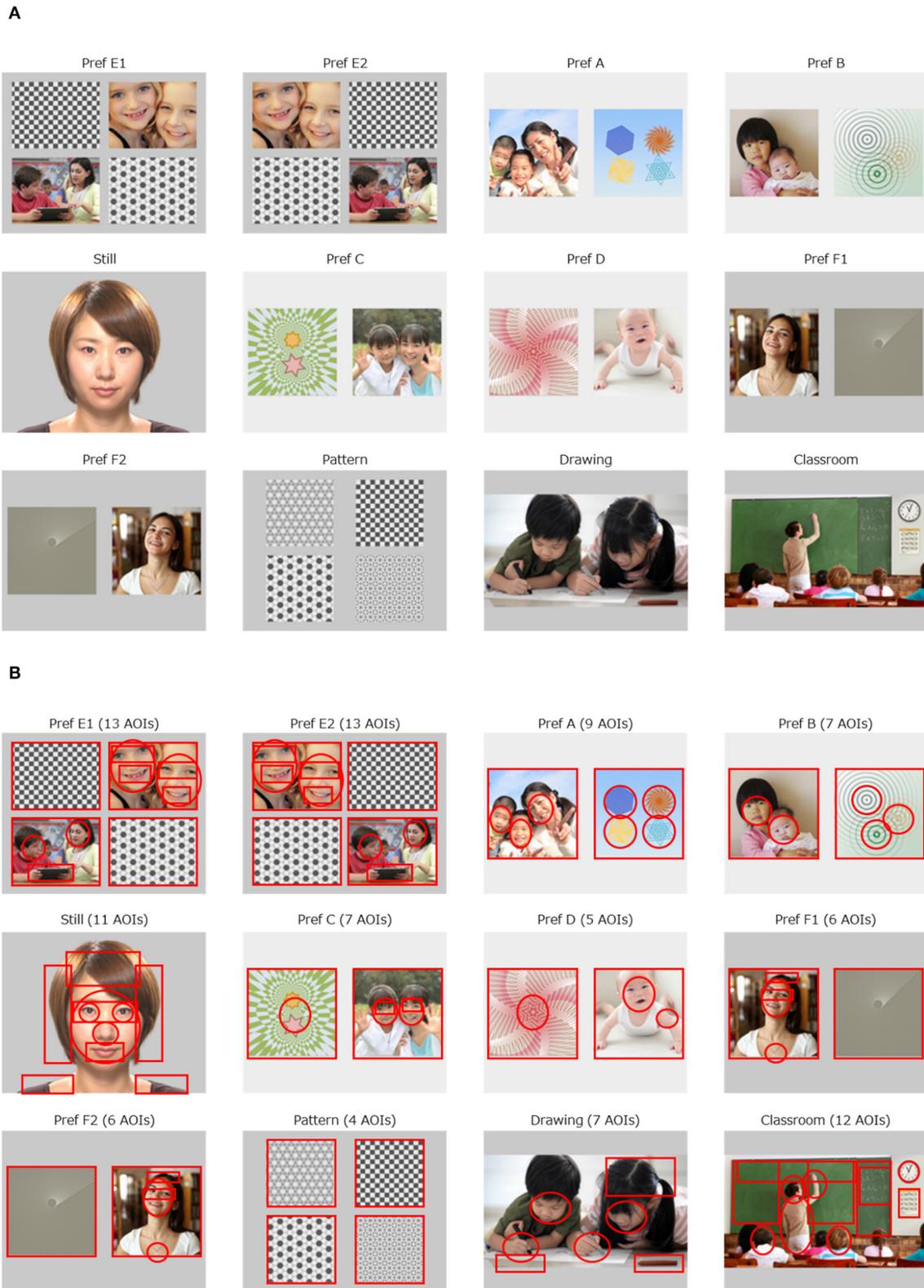


FIGURE 1 | Movie clips implemented with Gazefinder, and the AOIs. **(A)** The 12 movie clips. **(B)** 100 AOIs embedded in 12 movie clips.

still face of a young amateur actress (“Still,” 4.5 s), two children drawing pictures cooperatively (“Drawing,” 7.0 s), and a teacher teaching a math class in a classroom (“Classroom,” 11.0 s). The preferential paradigm was represented in nine movie clips, in which the visual field on the monitor was divided into two or four regions of equal size, with human figures or objects in parallel. The preferential paradigm movie clips consisted of one movie clip of four graphical patterns (“Pattern,” 7.0 s), and of eight movie clips with human figures with geometric patterns aligned side by side (“Pref A,” 5.0 s, “Pref B,” 4.5 s; “Pref C,” 5.0 s, “Pref D,” 4.5 s, “Pref E1,” 5.0 s, “Pref E2,” 4.5 s, “Pref F1,” 6.0 s, “Pref F2,” 5.5 s). The duration of these movie clips was set identical to that used in our previous studies. The remaining movie clips were kept as short as possible while preserving their context.

The rationale for creating the movie clips was as follows. As for the social paradigm, there is sufficient evidence to support the fact that a decreased gaze fixation on human faces, especially on eye regions on the monitor is associated with a diagnosis of ASD, regardless of age or IQ (14, 29). We have previously reported that fixation onto the eye region decreases in individuals with ASD at ages of 10 years and older (18). In addition, a decreased amount of gaze fixation onto human figures in social scenes has been reported consistently in individuals with ASD (14). Furthermore, it has been suggested that the quantity of the social content (e.g., the number of human figures presented in the scene) as well as its quality (e.g., human interaction) is related to the amount of reduction in gaze fixation in individuals with ASD (14, 15). Thus, the duration of gaze fixation is measured to contribute to the likelihood of ASD diagnosis, probably in accordance with the number of human figures (2 in “Drawing” and 8 in “Classroom”). Therefore, we presented two children drawing pictures while interacting in the movie clip “Drawing”. As for the preferential paradigm, we have previously confirmed that human figures were more preferentially looked at than non-social figures in typically developing individuals compared to individuals with ASD (18). To control the spatial preference (e.g., adherence to the right half of the visual field) that may be present in some participants with ASD (30), two sets of movie clips were duplicated (“Pref E1” vs. “Pref E2,” “Pref F1” vs. “Pref F2”), but the allocation of the targets (human figures) were exchanged horizontally, and inserted as different movie clips. In the other four sets of movies, the appearance of the side (left or right) of the target (human figures) was balanced.

Quantification of Eye Gaze Patterns

Through the sequence of the 12 movie clips, we defined 100 areas of interest (AOIs) in circles or squares on each movie clip, specified with x and y axes on the monitor (Figure 1B). We calculated two types of eye gaze indices: the AOI rate score and the AOI count score. The AOI rate score was defined as the percentage of gaze fixation time allocated to each AOI divided by the duration of each movie clip. The AOI rate score was between 0.0 and 1.0 and represents the focus on the object in a dose-response manner (i.e., the higher the AOI rate score, the more intensively the child focused on the object). The AOI count score was a representation of the presence (or absence) of a fixed gaze over each AOI, regardless of the duration of the

eye gaze. The possible AOI count score was 0 or 1 and reflects the presence of eye gaze on the AOI, irrespective of possible distractions occurring due to the child focusing on another AOI because of knowledge-driven prediction (31). This is likely to emerge among older individuals. This distraction also occurs when a human agent on the monitor acts as if she/he looked at the individual in front of the monitor (32). As such, we expected that the AOI count score would be more suitable for older individuals. We calculated both the AOI rate scores and AOI count scores separately for all 100 AOIs. In addition, we also calculated two different methods for AOI rate scores applied to the 100 AOIs. The first is to calculate the AOI rate scores of the first 1.0 s and the other is to calculate the AOI rate scores of the first 2.0 s. The intention for this was to generate more attributes with diagnostic value. Specifically, young individuals with ASD have been reported to pay less attention to faces during the initial viewing period (33). As a result, 300 sets of calculation for AOI rate scores and 100 sets of calculation for AOI count scores were applied.

Analysis

Strategy for Creating and Validating the Diagnostic Algorithm

Selection of Participants

We selected participants to generate a training dataset to realize the computerized diagnostic algorithm, and an independent dataset for the validation. We first excluded 57 participants from the following analyses because they had a diagnosis of ADHD ($N = 29$), general cognitive delay ($N = 24$), or a clinical diagnosis of epilepsy ($N = 4$). The intention was to minimize the neurophysiological heterogeneity of the subjects included in the dataset, except for the difference between ASD and typical development (TD). The remaining 165 participants were divided into three groups: ASD, TD, and the second control group. The ASD group ($N = 39$) consisted of individuals with a diagnosis of ASD confirmed with ADOS-2 or ADI-R and with a negative screening result for ADHD. The TD group ($N = 102$) consisted of individuals fulfilling negative screening results for both ASD and ADHD. The ASD and TD groups were primarily used as the source of the best-fit computerized diagnostic algorithm and the training and test datasets to check the validity of the best-fit diagnostic algorithm. The second control group consisted of two types of individuals: (1) those with a diagnosis of ASD with a positive screening result for ADHD, and (2) those without a confirmed diagnosis of ASD but with a positive screening result for ASD (the screening result for ADHD can be either positive or negative). The second control group served as an independent sample to validate the diagnostic predictability of the best-fit diagnostic algorithm.

We further divided the ASD and TD groups according to age. Although the social paradigm was assumed to be age-independent (14), the preferential paradigm was reported to distinguish ASD from TD individuals, particularly when the subjects were 10 years and older (18). Considering these, both the ASD and TD groups were divided into younger (<10 years) and older (10 years and older) groups, respectively. We set the age of 10 as the breaking point as in the previous study (18) and

because of the statistical reason that 10 was the median age of the individuals in the ASD and TD groups.

Extracting Indices (Candidate Attributes)

We calculated the 300 sets of AOI rate scores and 100 sets of AOI count scores for all participating individuals. To extract indices to be included in the best-estimate diagnostic algorithm, the mean values for both AOI rate scores and AOI count scores were compared between the ASD and TD groups for the younger and older age bands. When we found indices that were significantly ($p < 0.05$) associated with the ASD diagnosis or had an effect size (Cohen's d) of 0.5 or larger, we retained these as *candidate attributes*, the indices to be included in the next step. To this end, we extracted four sets of candidate attributes based on the AOI rate scores and count scores of the younger and older individuals. To minimize unnecessary weights and to avoid overfitting resulting from choosing multiple AOIs out of one region on a movie clip, we chose only one attribute with the largest effect size out of the candidate attributes that were

calculated on the same AOI. This rule also applies to the three AOI rate scores that share the same AOI (the AOI rate scores calculated from the first 1.0 s, from the first 2.0 s, and from the beginning to the end of the movie clip). In addition, to avoid including chance findings with large effect size, we dropped candidate attributes that were extracted from the AOIs with gaze fixation percentage of $<20\%$, which corresponds roughly to a duration of 1.0 s or longer for most movie clips.

Creating the Best-Fit Diagnostic Algorithm

We first created four diagnostic algorithms, Algo #1 (the AOI rate scores for the younger individuals), #2 (the AOI rate scores for the older individuals), #3 (the AOI count scores for the younger individuals), and #4 (the AOI count scores for the older individuals), based on the four sets of candidate attributes combined (Figure 2). For each algorithm, the candidate attributes were either divided by the standard deviation, or dichotomized to 0 or 1 and summed, followed by a division by the number of the candidate attributes. These

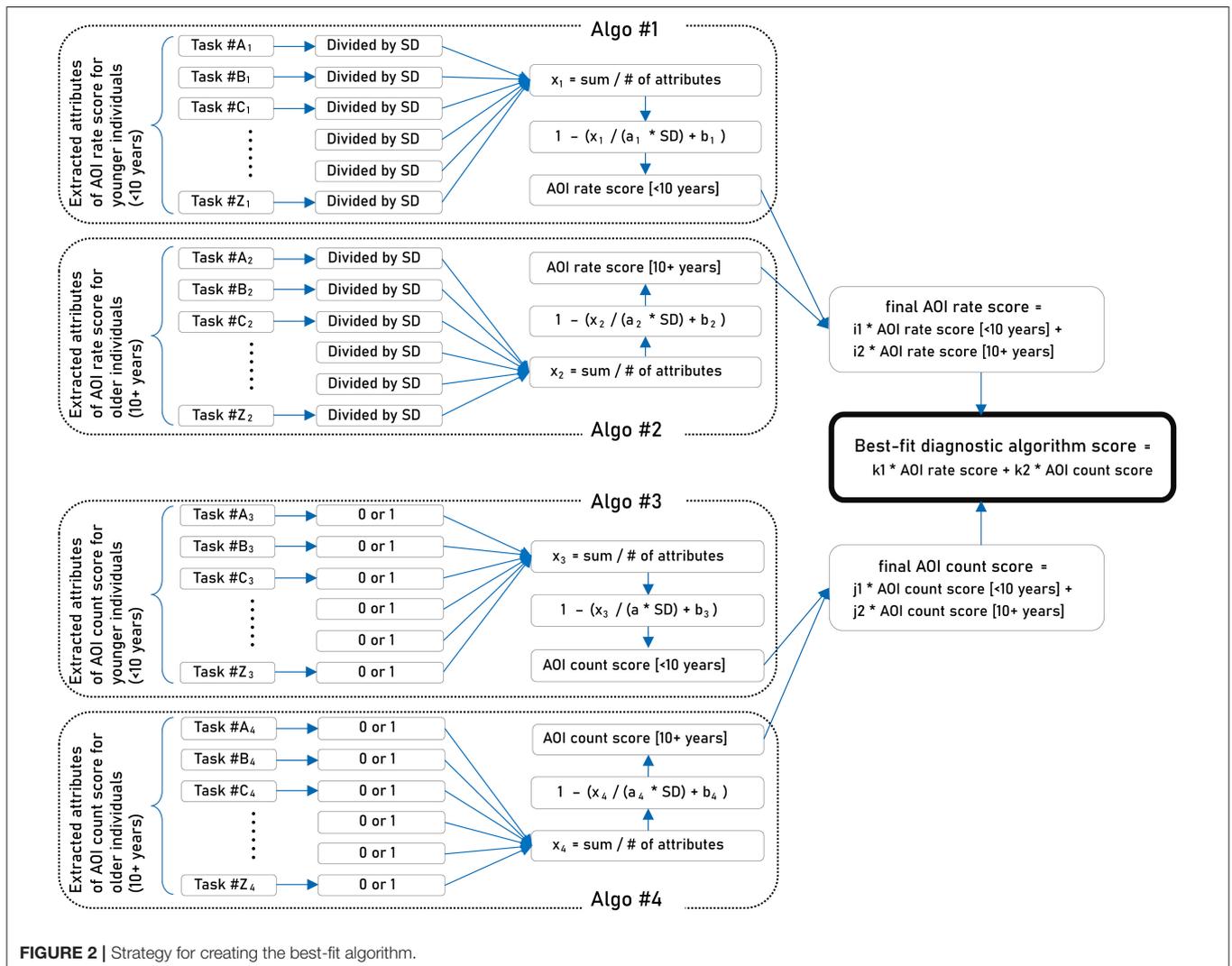


FIGURE 2 | Strategy for creating the best-fit algorithm.

values were fit for a sigmoid function that only took a value between 0 and 1. The next step was to merge Algo #1 and #2 to estimate the final AOI rate score, and to merge Algo #3 and #4 to estimate the final AOI count score. To merge two sigmoid functions, coefficients were estimated automatically to maximize the predictability of ASD diagnosis.

Next, we finalized the best-fit diagnostic algorithm using the two separately estimated algorithm-based scores: the final AOI rate score and the final AOI count score. Before merging the two scores, we chose one algorithm score of a better fit in the younger and older participants separately, to maximize the predictive validity. The fit was assessed with the area under the receiver operating characteristic (ROC) curve (AUC) for each set of comparisons. We then made the selected estimated scores merge smoothly as a continuum along the age range of the participants of 5–17 years (the best-fit algorithm). To merge the two sigmoid functions, coefficients were again estimated automatically to maximize the predictability of the ASD diagnosis. The best-fit algorithm score was made to take values between 0 and 1. In the following analyses, a value of 0.5 or higher of the best-fit algorithm score was assumed as an indicator of the individual under investigation having ASD.

Evaluation of the Diagnostic Performance (Cross-Validation)

We evaluated the general classification performance of the best-fit algorithm using the leave-one-out (LOO) method (34). We repeated the procedures described above, including extraction of candidate attributes and formulation of four separate algorithms to be merged into a single best-fit diagnostic algorithm, without the inclusion of one specific individual (LOO algorithm). The removed individual was tested for ASD based on the LOO

algorithm. This procedure was iterated for all participating individuals ($N = 141$; cross-validation). We then drew the ROC curves and calculated AUCs for both the best-fit diagnostic algorithm, and votes of the 141 LOO algorithms were used to interpret whether the validity of the best-fit algorithm might have been compromised. To simplify the interpretation, we also calculated the sensitivity, specificity, and accuracy for the best-fit algorithm and for the votes of 141 LOO algorithms separately. The point where the sensitivity and specificity were extracted was set at the Youden J index (i.e., sensitivity + specificity – 1) was maximized (35).

Evaluation of the Diagnostic Performance in an Independent Sample

Using the second control group ($N = 24$), we checked the diagnostic validity of the best-fit algorithm independently. Since this group consisted of individuals with ASD with a positive screening result for ADHD ($N = 17$) and individuals with no diagnosis of ASD but with a positive screening result for ASD ($N = 7$), we applied the best-fit algorithm and checked whether the diagnosis predicted with the best-fit algorithm score complied with the real diagnosis using AUC, sensitivity, specificity, and accuracy.

Statistics

All statistical analyses were conducted using Stata version 15.1 and R version 3.6.2. To calculate AUC values with 95% confidence intervals, ROC-kit 0.91 was used for resampling. For comparison of two continuous variables, we carried out either the t -test or the Kruskal-Wallis test, depending on the distribution. To avoid missing any potential candidate attributes at the early stage of the analyses, we set 0.05 as the significance level.

TABLE 1 | Characteristics of the participants.

	TD	ASD	Second control	Statistics*
Number of subjects	102	39	24	
Age in years, mean (SD)	9.5 (4.0)	10.3 (4.0)	10.4 (3.6)	$F_{(2,162)} = 0.86, p = 0.42$
Male sex, number (%)	43 (42%)	30 (77%)	15 (63%)	$\chi^2_{(2)} = 14.6, p = 0.001$
IQ/DQ, mean (SD)	104.1 (13.6)	94.5 (12.3)	98.7 (14.4)	$F_{(2,162)} = 7.64, p < 0.001$ TD > ASD
ASD screening [PARS total], mean (SD)	0.8 (1.1)	7.8 (5.2)	8.0 (7.0)	$F_{(2,162)} = 70.6, p < 0.001$ TD < ASD, TD < Second control
ADHD screening [ADHD-RS inattention], mean (SD)	3.3 (3.5)	7.9 (4.8)	11.7 (6.7)	$F_{(2,162)} = 42.1, p < 0.001$ TD < ASD, TD < Second control ASD < Second control
ADHD screening [ADHD-RS hyperactivity], mean (SD)	1.9 (3.0)	5.2 (4.4)	7.9 (6.2)	$F_{(2,162)} = 26.6, p < 0.001$ TD < ASD, TD < Second control ASD < Second control
Diagnosed as having ASD, number (%)	0 (0%)	39 (100%)	17 (71%)	$\chi^2_{(2)} = 142.9, p < 0.001$
Overall gaze fixation percentage, mean (SD)	92.1 (7.3)	89.0 (10.2)	91.2 (8.1)	$F_{(2,162)} = 2.24, p = 0.11$

*Statistically significant results after one-way ANOVA (F tests) were followed by group comparison with Bonferroni correction.

RESULTS

Characteristics of the Participating Individuals

Table 1 shows the demographic and clinical characteristics of the participants. There was no significant difference in the mean age of the participants among the groups. Compared with the TD group, the ASD group showed significantly lower IQ and higher scores on the ASD screening scale (PARS total) and ADHD screening scales (inattention and hyperactivity subscales of ADHD-RS).

Overall Gaze Fixation Percentage (Success of Data Retrieval)

The bottom row of **Table 1** shows that the overall gaze fixation percentage values during the measurement using Gazefinder were not statistically different across the groups (92% in the TD group, 89% in the ASD group, 91% in the second control group). The lowest value was 47.4% of a child belonging to the ASD group, but this was the only record below 60%. Out of the 165 participants, 161 showed a value of 70% or higher.

Extracting Indices (Candidate Attributes)

As for the first steps to create the best-fit diagnostic algorithm, we extracted the four sets of the candidate attributes to be used in the algorithm. The candidate attributes are shown on the corresponding AOIs in association with the AOI rate scores in **Supplementary Figures 1, 2** for the younger and older individuals, respectively, and in association with the AOI count scores in **Supplementary Figures 3, 4** in the younger and older individuals, respectively.

TABLE 2 | Calculated values of the area under curve (AUC) and their 95% confidence intervals (CIs) for the proposed algorithms for younger and older participants using either AOI rate scores or AOI count scores.

	Subjects to be tested	AUC	95% CI
Algo #1: AOI rate score for the younger individuals	Younger individuals ($n = 80$)	0.83	0.72–0.90
Algo #2: AOI rate score for the older individuals	Older individuals ($n = 61$)	0.83	0.72–0.92
Final AOI rate score algorithm	Younger individuals ($n = 80$)	0.82	0.70–0.90
Final AOI rate score algorithm	Older individuals ($n = 61$)	0.82	0.69–0.92
Algo #3: AOI count score for the younger individuals	Younger individuals ($n = 80$)	0.74	0.63–0.83
Algo #4: AOI count score for the older individuals	Older individuals ($n = 61$)	0.88	0.78–0.95
Final AOI count score algorithm	Younger individuals ($n = 80$)	0.75	0.63–0.85
Final AOI count score algorithm	Older individuals ($n = 61$)	0.87	0.72–0.95

Creating the Best-Fit Diagnostic Algorithm

Table 2 shows the AUCs for Algo #1, 2, 3, 4, the final AOI rate score algorithms, and the final AOI count score algorithms. To select one algorithm for the two age bands, we found that the final AOI rate score algorithm fit equally for the younger and older individuals (0.82 vs. 0.82) and that the final AOI count score algorithm showed a better fit for the older individuals (0.75 vs. 0.87). For the younger individuals, we selected the final AOI rate score algorithm, and for the older individuals, we selected the final AOI count score algorithm. Merging the two algorithms along the age bands provided the best-fit diagnostic algorithm, of which the ROC curve is shown as a solid line in **Figure 3**, and the AUC was 0.84, as shown in the first row of **Table 3**. The sensitivity, specificity, and accuracy were 74, 80, and 78%, respectively. We also checked whether the best-fit diagnostic algorithm showed a good fit for the younger and older participants. The second and third rows of **Table 3** show that the AUC, sensitivity, specificity, and accuracy did not differ between the younger and older participants.

Evaluation of Diagnostic Performance

To show that the high accuracy of the best-fit diagnostic algorithm did not result from overfitting or from chance alone, the diagnostic performance of the best-fit algorithm was first

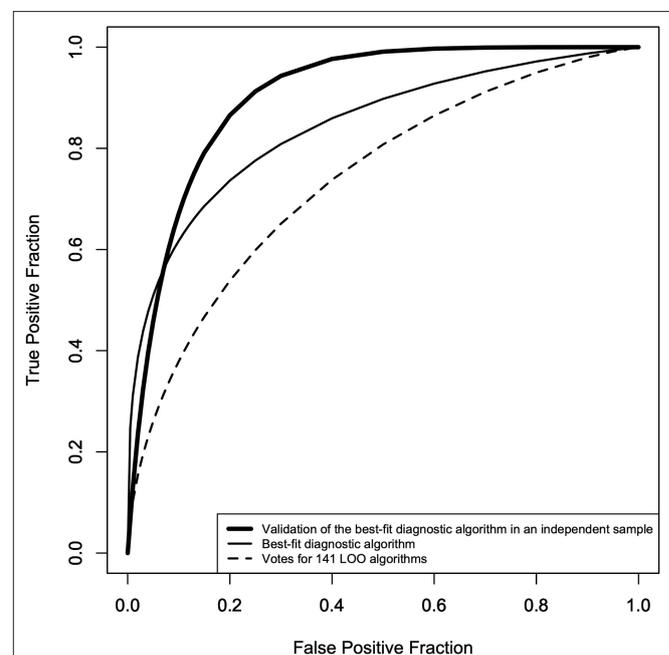


FIGURE 3 | ROC curves of the best-fit diagnostic algorithm* and the votes of the 141 LOO** algorithms. Solid line: ROC curve of the best-fit diagnostic algorithm (AUC = 0.84, sensitivity = 71%, specificity = 80%, accuracy = 78%), dotted line: ROC curve of the votes of the 141 LOO algorithms (AUC = 0.74, sensitivity = 65%, specificity = 70%, accuracy = 67%), thick line: ROC curve of the best-fit diagnostic algorithm in an independent sample (second control group: AUC = 0.91, sensitivity = 87%, specificity = 80%, accuracy = 88%). *The merged algorithm of the final AOI rate score algorithm for age <10 years, and the final AOI count score algorithm for 10 years and over. **Leave-one-out method to cross-validate the best-fit diagnostic algorithm.

assessed using the LOO method, the result of which is shown as a dotted line in **Figure 3** and in the fourth row of **Table 3**. The AUC was 0.74, which was smaller than the AUC of the best-fit algorithm, and the sensitivity and specificity were 65 and 70%, respectively. We also checked the diagnostic performance of the best-fit diagnostic algorithm in an independent sample (the second control group, $N = 24$). For the 24 participants in this group, we found an AUC of 0.91, with a sensitivity and specificity of 87 and 80%, respectively.

DISCUSSION

Using Gazefinder, we successfully created the best-fit diagnostic algorithm to discriminate school-aged and adolescent individuals with ASD from typically developing individuals of the same age range, with a sensitivity of 78% and specificity of 80%. The diagnostic performance was tested in two ways: one was a machine-learning procedure called the LOO method and the other was a test in a different, independent sample of the same age range. These two tests of diagnostic performance indicated acceptable to excellent discriminability.

The reported sensitivity and specificity of ADOS-2 (Module 3) were as high as 91 and 66%, respectively, in a large sample of German children with an average age of 10 (36). Similarly, the sensitivity and specificity of ADI-R in a small sample of Japanese children of age 5–9 were 92 and 84%, respectively, and of age 10–19 were 97 and 90%, respectively (9). The diagnostic accuracy of Gazefinder was not better than that of the standard diagnostic tools, but comparable. In addition, the diagnostic validity of Gazefinder was even better than the established screening tools that are available for a wide range of ages. For instance, the Social Communication Questionnaire (37) is a widely used tool available for a wide age range, although the sensitivity and specificity were 64 and 72%, respectively, in a sample of 1–25-year-old individuals (38). There are also a number of screening tools with reported sensitivities and specificities exceeding 80%, although these figures have not been cross-validated or tested in different datasets (39). Considering the applicability of the best-fit algorithm to a wide age range of the subjects, our validated data support the use of the best-fit algorithm in clinical settings as an available alternative to a range of screening tools for detecting ASD, particularly in terms of diagnostic performance.

In addition to the fact that the diagnostic evaluation using Gazefinder was completed within 2 min without any expertise, it is worth noting that a substantial majority of the participating individuals succeeded in completing the examination. Surprisingly, 161 (98%) of the 165 participants kept staring at the monitor for 70% or longer of the total duration of the examination. The four individuals with <70% of the total fixation time were diagnosed as having ASD. Among these four individuals, only two individuals were correctly predicted as having ASD (data not shown in tables). Apparently, the diagnostic accuracy may be decreased if we include individuals with <70% of the overall gaze fixation percentage. One explanation for this observation is that the lower overall gaze fixation percentage by itself may be predictive of ASD. This leads to an understanding that the accuracy of the best-fit algorithm might have been compromised when the overall gaze fixation percentage was lower than the specific cutoff, for example, 70%. Although the overall accuracy of the best-fit diagnostic algorithm was secured, we may set a threshold of 70% as the lowest percentage for securing algorithm-based diagnosis for clinical use, until firm conclusions are drawn.

Discussion of the Methodology

There was an initial possibility of potential overfitting due to the limited sample size in our results. This has been discussed in the context where the attributes outnumber the observations in machine-learning-assisted neuroimaging studies (40). We have made several attempts to overcome this shortcoming. First, to enhance the efficiency in creating a valid algorithm, we tried to increase the clinical homogeneity among the diagnostic groups. We extracted participants with ASD without any comorbid conditions and TD participants without any suspicions of ASD symptomatology. Second, we avoided building a multi-layered algorithm. Until recently, neural networks, and their applications such as in deep learning, the prominent feature of which is a combination of perceptrons aligned in multiple layers, has been used widely in literature. The major problem inherent to these techniques is overfitting, particularly if the sample size is small, when the network cannot learn itself (41). Therefore, we adopted a single-layered algorithm. Third, we adopted cross-validation using the LOO method (34). Cross-validation is required not only for checking the predictive validity, but also for achieving optimal

TABLE 3 | Diagnostic performance of the best-fit algorithm.

	Subjects to be tested	AUC	95%CI	Sensitivity**	Specificity**	Accuracy**
Best-fit algorithm	All ($n = 141$)	0.84	0.76–0.91	0.74	0.80	0.78
Best-fit algorithm	Younger individuals ($n = 80$)	0.82	0.70–0.90	0.78	0.70	0.76
Best-fit algorithm	Older individuals ($n = 61$)	0.87	0.72–0.96	0.73	0.87	0.75
Votes for 141 LOO algorithms*	All ($n = 141$)	0.74	0.64–0.82	0.65	0.70	0.67
Best-fit algorithm	Second control group ($n = 24$)	0.91	0.66–0.99	0.87	0.80	0.88

*Leave-one-out (LOO) algorithm: an algorithm developed with a procedure identical to that applied to develop the best-fit algorithm without the inclusion of one specific individual (LOO algorithms). The removed individual was tested for the diagnosis of ASD, based on each LOO algorithm. This procedure was iterated for all participating individuals, and the majority of the votes for the 141 LOO algorithm was used as the cross-validated predicted diagnosis ($N = 141$; cross-validation).

**Sensitivity and specificity were calculated at the point on the ROC where the Youden J index (Sensitivity + Specificity – 1) was maximized.

diagnostic performance (42). LOO is assumed to perform better than other cross-validation methods because the test data are secured not to be used in the training data to form an algorithm. Furthermore, we used a different, independent dataset (the second control group) to be tested with the best-fit diagnostic algorithm. It is worth noting that the independent second control group was a mixture of individuals with ASD with clinical signs of ADHD, and non-TD individuals with subthreshold signs of either ASD or ADHD or both. However, the diagnostic performance of this sample was not compromised. To this end, our validation processes have supported the robust predictive validity of the proposed best-fit diagnostic algorithm.

Limitations

Despite the fact that the predictive validity of the best-fit diagnostic algorithm was established, potential limitations of the findings should be acknowledged. First, we enrolled a relatively small sample of individuals of Japanese origin. Our stimulus movie was created to be used among non-Japanese people as well, and included actors of various ethnicities. Our findings may be better replicated in a larger sample with different cultural and biological settings. Second, we invented the diagnostic algorithm based on responsivity to social stimuli; however, this is only one aspect of the broad behavioral spectrum of ASD. Particularly, we have not established that our measure reflects the symptoms of restricted interest and repetitive behaviors (RRBs). Furthermore, the predicted diagnosis does not indicate the severity of the symptoms, as the diagnostic algorithm has been designed to monitor whether responsivity to specific stimuli was observed or not. Thus, Gazefinder has immense possibility for customization in the future. Third, we did not investigate whether the indices we collected were associated with clinical correlates, severity, or prognosis. In a previous study using eye-tracking devices, children with ASD who were more oriented toward social images were shown to have better language and higher IQ scores (16). The clinical applicability of Gazefinder can be further developed in this direction in the context of treatment monitoring. Fourth, we did not assess social anxiety symptoms. In a previous study, gaze avoidance was reported in adolescents with either social anxiety disorder or ASD, but delayed orientation to the eye regions was observed in the latter. We did not examine whether the reduced gaze fixation to the eye regions results from delayed orientation or from orienting in a direction outside the eye regions; this should be addressed in the future. Fifth, we did not assess the participants of our study with both ADOS-2 and ADI-R-JV; we assessed them with only one of these tools. Among 39 individuals with ASD in the analysis, five were assessed only with ADI-R, and two individuals among these were over 10 years. This may compromise the diagnostic accuracy because of higher likelihood of recall biases, although the number of such individuals is minimal. Sixth, we excluded individuals with ADHD and general cognitive delay. Although this exclusion will allow the algorithm to be more sensitive to ASD, the general clinical applicability of the diagnostic algorithm may be limited in clinical settings, where individuals with ASD are frequently comorbid with ADHD and/or cognitive delay. In our future study, we may include individuals with or without

ADHD and compare them with individuals with ASD using the diagnostic algorithm. Finally, since we did not conduct full diagnostic assessments for screen-negative individuals, we might have overlooked ASD diagnoses in this group of individuals. However, this is unlikely since our thorough clinical assessments were conducted by trained psychiatrists or pediatricians, all of whom have an experience of joining clinical/research workshop of ADOS-G or ADOS-2 and some have established research reliability with the developers, followed by the consistent negative results for all the three screening tests for ASD.

Clinical Applicability

In typical community settings, individuals with ASD are expected to be diagnosed at certain stages during childhood (43). However, more than half of the children, adolescents, and young adults with confirmed diagnosis of ASD do not have a history of clinical diagnoses related to ASD, as was reported in a community survey in the last century from the US (44). This was reported a decade ago, although the situation appears to remain the same at present. A more recent study from Japan pointed out that only 32% of children confirmed to have ASD at 5 years of age had any history of neuropsychiatric/neuropediatric diagnosis until their fifth birthdays, meaning that more than half of the children with ASD are left undiagnosed at 5 years of age or even older (1). This may be due to the lack of appropriate chances to be screened, although general health checkups during childhood are a rule in most developed countries. The challenges inherent to the diagnostic evaluation of ASD, particularly in the community setting, should be resolved with ease and without expertise. We propose the computerized best-fit diagnostic algorithm implemented in Gazefinder as a solution to this.

Despite the diagnostic accuracy and convenience of the diagnostic algorithm, standard diagnostic procedures in clinical settings, such as ADOS-2, should not be replaced with diagnosis with Gazefinder. We assume that a diagnostic evaluation with Gazefinder is a mechanical one and thus should be followed with an expert clinical diagnosis. Presently, clinical evaluation may not be readily available in countries where trained manpower is limited, and mechanical diagnosis alone can result in a false sense of security among caregivers. However, in such countries, we can propound that Gazefinder can function as a screener and thereby reduce the burden on experts including pediatricians, child psychologists, and child psychiatrists. We should minimize the drawbacks and maximize the advantages of using Gazefinder in the future. In order not to give false sense of security when a false negative result is provided, the cutoff point should be adjusted to increase the sensitivity.

We have confirmed that the diagnostic performance of the best-fit algorithm is comparable to standard diagnostic tools and is even better than current screening tools for ASD. Diagnostic evaluation using Gazefinder is secured in more than 90% of the participants and adolescents. Thus, the proposed best-fit diagnostic algorithm is ready to be used in clinical settings and to be tested in clinical trials. We have drafted and submitted the protocol to the Japanese supervisory authorities, and currently, a clinical trial is under way. Clinical trials using Gazefinder to

establish diagnostic validity in countries other than Japan are also appreciated.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because sharing the data with the third parties not listed in the original protocol has not been approved by the Ethical Committees. Requests to access the datasets should be directed to Kenji J. Tsuchiya, tsuchiya@hama-med.ac.jp.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Hamamatsu University School of Medicine and Hospital, the Ethics Committee of University of Fukui, the Ethics Committee of Hirosaki University School of Medicine and Hospital, the Ethics Committee of Graduate School of Medicine and School of Medicine, Chiba University, the Ethics Committee of Saga Medical School Faculty of Medicine, Saga University, the Ethics Committee of Kanazawa University School of Medicine, and the Ethics Committee of Tottori University Faculty of Medicine. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

KT, SH, THara, MN, MS, HK, YH, MM, MK, YM, and TK: study concept and design. KT, MS, TF, HK, YH, MM, MK, YM, and THarada: resources arrangement, clinical evaluation, and measurement. KT, SH, THara, MN, and TN: analysis and interpretation of data. KT, THara, and TN: drafting of manuscript. KT and TK: obtained funding. All authors contributed to critical revision of manuscript for important intellectual content and approved the submitted version.

FUNDING

This study was supported by AMED under Grant Numbers 18he1102012h0004 and 20hk0102064h0002.

ACKNOWLEDGMENTS

Gazefinder is the trademark or a registered trademark of JVCKENWOOD Corporation. The authors thank all the study

participants for their voluntary cooperation in this study. The authors also thank Ms. Noriko Kodera, Emi Higashimoto, Chikako Nakayasu, Yuko Amma, Haruka Suzuki, Miho Otaki, Taeko Ebihara, Chikako Fukuda, Mr. Katsuyuki Shudo, Hirokazu Sekiguchi, Shunichi Takeichi, Drs. Tomoko Haramaki, Damee Choi, Nagahide Takahashi, Akemi Okumura, Toshiki Iwabuchi, Katsuaki Suzuki, Chie Shimmura, Jun Inoue, Hitoshi Kuwabara, Yoshifumi Mizuno, Shinichiro Takiguchi, Tokiko Yoshida, Aki Tsuchiyagaito, Tsuyoshi Sasaki, Ayumi Okato, Prof. Yoshinobu Ebisawa, Prof. Shinsei Minoshima, Prof. Norio Mori, Prof. Hidenori Yamasue, Prof. Hideo Matsuzaki, Prof. Akemi Tomoda, and Prof. Eiji Shimizu for data collection, helpful comments, and technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2020.603085/full#supplementary-material>

Supplementary Figure 1 | Candidate attributes* of AOI rate scores** among the individuals of younger than 10 years. AOIs outlined with red lines represent parameters with a positive effect (i.e., AOI rate score was higher in individuals with ASD) and AOIs outlines with blue lines represent parameters with a negative effect (i.e., AOI rate score was lower in individuals with ASD). *Attributes that were significantly ($p < 0.05$) associated with the diagnosis of ASD or had an effect size of Cohen's d with 0.5 or larger. **Percentage of fixation time allocated to the AOI divided by duration of each movie clip.

Supplementary Figure 2 | Candidate attributes* of AOI rate scores** among the individuals of 10 years and over. AOIs outlined with red lines represent parameters with a positive effect (i.e., AOI rate score was higher in individuals with ASD) and AOIs outlines with blue lines represent parameters with a negative effect (i.e., AOI rate score was lower in individuals with ASD). *Attributes that were significantly ($p < 0.05$) associated with the diagnosis of ASD or had an effect size of Cohen's d with 0.5 or larger. **Percentage of fixation time allocated to the AOI divided by duration of each movie clip.

Supplementary Figure 3 | Candidate attributes* of AOI count scores** among the individuals of younger than 10 years. AOIs outlined with red lines represent parameters with a positive effect (i.e., AOI count score was higher in individuals with ASD) and AOIs outlines with blue lines represent parameters with a negative effect (i.e., AOI count score was lower in individuals with ASD). *Attributes that were significantly ($p < 0.05$) associated with the diagnosis of ASD or had an effect size of Cohen's d with 0.5 or larger. **Presence (or absence) of countinous fixed gaze over the AOI; takes the value of either 0 or 1.

Supplementary Figure 4 | Candidate attributes* of AOI count scores** among the individuals of 10 years and over. AOIs outlined with red lines represent parameters with a positive effect (i.e., AOI count score was higher in individuals with ASD) and AOIs outlines with blue lines represent parameters with a negative effect (i.e., AOI count score was lower in individuals with ASD). *Attributes that were significantly ($p < 0.05$) associated with the diagnosis of ASD or had an effect size of Cohen's d with 0.5 or larger. **Presence (or absence) of countinous fixed gaze over the AOI; takes the value of either 0 or 1.

REFERENCES

- Saito M, Hirota T, Sakamoto Y, Adachi M, Takahashi M, Osato-Kaneda A, et al. Prevalence and cumulative incidence of autism spectrum disorders and the patterns of co-occurring neurodevelopmental disorders in a total population sample of 5-year-old children. *Mol Autism*. (2020) 11:35. doi: 10.1186/s13229-020-00342-5
- van Heijst BE, Geurts HM. Quality of life in autism across the lifespan: a meta-analysis. *Autism*. (2015) 19:158–67. doi: 10.1177/1362361313517053
- Wong C, Odom SL, Hume KA, Cox AW, Fettig A, Kucharczyk S, et al. Evidence-based practices for children, youth, and young adults with autism

- spectrum disorder: a comprehensive review. *J Autism Dev Disord.* (2015) 45:1951–66. doi: 10.1007/s10803-014-2351-z
4. Clark MLE, Vinen Z, Barbaro J, Dissanayake C. School age outcomes of children diagnosed early and later with autism spectrum disorder. *J Autism Dev Disord.* (2018) 48:92–102. doi: 10.1007/s10803-017-3279-x
 5. Barbaro J, Dissanayake C. Prospective identification of autism spectrum disorders in infancy and toddlerhood using developmental surveillance: the social attention and communication study. *J Dev Behav Pediatr.* (2010) 31:376–85. doi: 10.1097/DBP.0b013e3181df7f3c
 6. Maenner MJ, Schieve LA, Rice CE, Cunniff C, Giarelli E, Kirby RS, et al. Frequency and pattern of documented diagnostic features and the age of autism identification. *J Am Acad Child Adolesc Psychiatry.* (2013) 52:401–13 e8. doi: 10.1016/j.jaac.2013.01.014
 7. Pierce K, Gazestani VH, Bacon E, Barnes CC, Cha D, Nalobolu S, et al. Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA pediatr.* (2019) 173:578–87. doi: 10.1001/jamapediatrics.2019.0624
 8. Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL. *ADOS-2 Autism Diagnostic Observation Schedule Second Edition.* Torrance, CA: Western Psychological Services (2012).
 9. Tsuchiya KJ, Matsumoto K, Yagi A, Inada N, Kuroda M, Inokuchi E, et al. Reliability and validity of autism diagnostic interview-revised, Japanese version. *J Autism Dev Disord.* (2013) 43:643–62. doi: 10.1007/s10803-012-1606-9
 10. Bent CA, Barbaro J, Dissanayake C. Parents' experiences of the service pathway to an autism diagnosis for their child: what predicts an early diagnosis in Australia? *Res Dev Disabil.* (2020) 103:103689. doi: 10.1016/j.ridd.2020.103689
 11. Zwaigenbaum L, Penner M. Autism spectrum disorder: advances in diagnosis and evaluation. *BMJ.* (2018) 361:k1674. doi: 10.1136/bmj.k1674
 12. Constantino JN, Kennon-McGill S, Weichselbaum C, Marrus N, Haider A, Glowinski AL, et al. Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature.* (2017) 547:340–4. doi: 10.1038/nature22999
 13. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5).* Washington, DC: American Psychiatric Publishing Group (2013).
 14. Frazier TW, Strauss M, Klingemier EW, Zetzer EE, Hardan AY, Eng C, et al. A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. *J Am Acad Child Adolesc Psychiatry.* (2017) 56:546–55. doi: 10.1016/j.jaac.2017.05.005
 15. Chita-Tegmark M. Social attention in ASD: a review and meta-analysis of eye-tracking studies. *Res Dev Disabil.* (2016) 48:79–93. doi: 10.1016/j.ridd.2015.10.011
 16. Pierce K, Marinero S, Hazin R, McKenna B, Barnes CC, Malige A. Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. *Biol Psychiatry.* (2016) 79:657–66. doi: 10.1016/j.biopsych.2015.03.032
 17. Moore A, Wozniak M, Yousef A, Barnes CC, Cha D, Courchesne E, et al. The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking. *Mol Autism.* (2018) 9:19. doi: 10.1186/s13229-018-0202-z
 18. Fujioka T, Tsuchiya KJ, Saito M, Hirano Y, Matsuo M, Kikuchi M, et al. Developmental changes in attention to social information from childhood to adolescence in autism spectrum disorders: a comparative study. *Mol Autism.* (2020) 11:24. doi: 10.1186/s13229-020-00321-w
 19. Fujioka T, Inohara K, Okamoto Y, Masuya Y, Ishitobi M, Saito DN, et al. Gazefinder as a clinical supplementary tool for discriminating between autism spectrum disorder and typical development in male adolescents and adults. *Mol Autism.* (2016) 7:19. doi: 10.1186/s13229-016-0083-y
 20. Adachi J, Yukihiro R, Inoue M, Tsujii M, Kurita H, Ichikawa H, et al. Reliability and validity of short version of pervasive developmental disorders autism society Japan rating scale (PARS): a behavior checklist for people with PDD (in Japanese). *Seishin Igaku.* (2008) 50:431–8. doi: 10.1037/t68535-000
 21. Goodman R. The strengths and difficulties questionnaire: a research note. *J Child Psychol Psychiatry.* (1997) 38:581–6. doi: 10.1111/j.1469-7610.1997.tb01545.x
 22. Kamio Y, Inada N, Moriwaki A, Kuroda M, Koyama T, Tsujii H, et al. Quantitative autistic traits ascertained in a national survey of 22 529 Japanese schoolchildren. *Acta Psychiatr Scand.* (2013) 128:45–53. doi: 10.1111/acps.12034
 23. DuPaul GJ, Power TJ, McGoey K, Ikeda M, Anastopoulos AD. Reliability and validity of parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms. *J Psychoeduc Assess.* (1998) 16:55–8. doi: 10.1177/073428299801600104
 24. Uno Y, Mizukami H, Ando M, Yukihiro R, Iwasaki Y, Ozaki N. Reliability and validity of the new Tanaka B Intelligence Scale scores: a group intelligence test. *PLoS ONE.* (2014) 9:e100262. doi: 10.1371/journal.pone.0100262
 25. Koyama T, Osada H, Tsujii H, Kurita H. Utility of the Kyoto Scale of Psychological Development in cognitive assessment of children with pervasive developmental disorders. *Psychiatry Clin Neurosci.* (2009) 63:241–3. doi: 10.1111/j.1440-1819.2009.01931.x
 26. Conners CK. *Conners 3rd Edition (in Japanese: Japanese translation by Y. Tanaka and R. Sakamoto).* Tokyo: Kanekoshobo (2011).
 27. Fujioka T, Fujisawa TX, Inohara K, Okamoto Y, Matsumura Y, Tsuchiya KJ, et al. Attenuated relationship between salivary oxytocin levels and attention to social information in adolescents and adults with autism spectrum disorder: a comparative study. *Ann Gen Psychiatry.* (2020) 19:38. doi: 10.1186/s12991-020-00287-2
 28. Yamasue H, Okada T, Munesue T, Kuroda M, Fujioka T, Uno Y, et al. Effect of intranasal oxytocin on the core social symptoms of autism spectrum disorder: a randomized clinical trial. *Mol Psychiatry.* (2020) 25:1849–58. doi: 10.1038/s41380-018-0097-2
 29. Chita-Tegmark M. Attention allocation in ASD: a review and meta-analysis of eye-tracking studies. *Rev J Autism Dev Disord.* (2016) 3:209–23. doi: 10.1007/s40489-016-0077-x
 30. Hessels RS. How does gaze to faces support face-to-face interaction? A review and perspective. *Psychon Bull Rev.* (2020) 27:856–81. doi: 10.31219/osf.io/8zta5
 31. Henderson JM. Gaze control as prediction. *Trends Cogn Sci.* (2017) 21:15–23. doi: 10.1016/j.tics.2016.11.003
 32. Colomatto C, van Buren B, Scholl BJ. Intentionally distracting: Working memory is disrupted by the perception of other agents attending to you - even without eye-gaze cues. *Psychon Bull Rev.* (2019) 26:951–7. doi: 10.3758/s13423-018-1530-x
 33. Amso D, Haas S, Tenenbaum E, Markant J, Sheinkopf SJ. Bottom-up attention orienting in young children with autism. *J Autism Dev Disord.* (2014) 44:664–73. doi: 10.1007/s10803-013-1925-5
 34. Li Q. Reliable evaluation of performance level for computer-aided diagnostic scheme. *Acad Radiol.* (2007) 14:985–91. doi: 10.1016/j.acra.2007.04.015
 35. Youden WJ. Index for rating diagnostic tests. *Cancer.* (1950) 3:32–5. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
 36. Medda JE, Cholemkery H, Freitag CM. Sensitivity and specificity of the ADOS-2 algorithm in a large German sample. *J Autism Dev Disord.* (2019) 49:750–61. doi: 10.1007/s10803-018-3750-3
 37. Rutter M, Bailey A, Lord C. *Social Communication Questionnaire.* Los Angeles, CA: Western Psychological Services (2003).
 38. Barnard-Brak L, Brewer A, Chesnut S, Richman D, Schaeffer AM. The sensitivity and specificity of the social communication questionnaire for autism spectrum with respect to age. *Autism Res.* (2016) 9:838–45. doi: 10.1002/aur.1584
 39. Thabtah F, Peebles D. Early autism screening: a comprehensive review. *Int J Environ Res Public Health.* (2019) 16:3502. doi: 10.3390/ijerph16183502
 40. Bajestani GS, Behrooz M, Khani AG, Nouri-Baygi M, Mollaei A. Diagnosis of autism spectrum disorder based on complex network features. *Comput Methods Programs Biomed.* (2019) 177:277–83. doi: 10.1016/j.cmpb.2019.06.006
 41. Sejnowski TJ. The unreasonable effectiveness of deep learning in artificial intelligence. *Proc Natl Acad Sci USA.* (2020) 117:30033–38. doi: 10.1073/pnas.1907373117
 42. Vabalas A, Gowen E, Poliakov E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS ONE.* (2019) 14:e0224365. doi: 10.1371/journal.pone.0224365
 43. Johnson CP, Myers SM, American Academy of Pediatrics Council on Children with Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics.* (2007) 120:1183–215. doi: 10.1542/peds.2007-2361
 44. Barbaresi WJ, Colligan RC, Weaver AL, Katusic SK. The incidence of clinically diagnosed versus research-identified autism in Olmsted County, Minnesota,

1976-1997: results from a retrospective, population-based study. *J Autism Dev Disord.* (2009) 39:464–70. doi: 10.1007/s10803-008-0645-8

Conflict of Interest: SH and MN are employed by JVCKENWOOD Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Tsuchiya, Hakoshima, Hara, Ninomiya, Saito, Fujioka, Kosaka, Hirano, Matsuo, Kikuchi, Maegaki, Harada, Nishimura and Katayama. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.