



## OPEN ACCESS

## EDITED BY

Wing Chee So,  
The Chinese University of Hong Kong,  
Hong Kong SAR, China

## REVIEWED BY

Nicole Wolff,  
Technische Universität, Germany  
Neelam Muthu Lakshmi,  
Sri Padmavati Mahila Visvavidyalayam, India

## \*CORRESPONDENCE

Ming Li  
✉ ming.li369@dukekunshan.edu.cn  
Xiao-bing Zou  
✉ zouxib@mail.sysu.edu.cn

<sup>†</sup>These authors have contributed equally to this work and share first authorship

## SPECIALTY SECTION

This article was submitted to  
Autism,  
a section of the journal  
Frontiers in Psychiatry

RECEIVED 08 September 2022

ACCEPTED 05 January 2023

PUBLISHED 26 January 2023

## CITATION

Zhu F-l, Wang S-h, Liu W-b, Zhu H-l, Li M and Zou X-b (2023) A multimodal machine learning system in early screening for toddlers with autism spectrum disorders based on the response to name.  
*Front. Psychiatry* 14:1039293.  
doi: 10.3389/fpsy.2023.1039293

## COPYRIGHT

© 2023 Zhu, Wang, Liu, Zhu, Li and Zou. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# A multimodal machine learning system in early screening for toddlers with autism spectrum disorders based on the response to name

Feng-lei Zhu<sup>1†</sup>, Shi-huan Wang<sup>1†</sup>, Wen-bo Liu<sup>2</sup>, Hui-lin Zhu<sup>1</sup>,  
Ming Li<sup>3\*</sup> and Xiao-bing Zou<sup>1\*</sup>

<sup>1</sup>Child Developmental and Behavioral Center, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, <sup>2</sup>School of Electronics and Information Technology, Guangzhou Higher Education Mega Center, Sun Yat-sen University, Guangzhou, China, <sup>3</sup>Data Science Research Center, Duke Kunshan University, Kunshan, China

**Background:** Reduced or absence of the response to name (RTN) has been widely reported as an early specific indicator for autism spectrum disorder (ASD), while few studies have quantified the RTN of toddlers with ASD in an automatic way. The present study aims to apply a multimodal machine learning system (MMLS) in early screening for toddlers with ASD based on the RTN.

**Methods:** A total of 125 toddlers were recruited, including ASD ( $n = 61$ ), developmental delay (DD,  $n = 31$ ), and typical developmental (TD,  $n = 33$ ). Procedures of RTN were, respectively, performed by the evaluator and caregiver. Behavioral data were collected by eight-definition tripod-mounted cameras and coded by the MMLS. Response score, response time, and response duration time were accurately calculated to evaluate RTN.

**Results:** Total accuracy of RTN scores rated by computers was 0.92. In both evaluator and caregiver procedures, toddlers with ASD had significant differences in response score, response time, and response duration time, compared to toddlers with DD and TD (all  $P$ -values  $< 0.05$ ). The area under the curve (AUC) was 0.81 for the computer-rated results, and the AUC was 0.91 for the human-rated results. The accuracy in the identification of ASD based on the computer- and human-rated results was, respectively, 74.8 and 82.9%. There was a significant difference between the AUC of the human-rated results and computer-rated results ( $Z = 2.71$ ,  $P$ -value = 0.007).

**Conclusion:** The multimodal machine learning system can accurately quantify behaviors in RTN procedures and may effectively distinguish toddlers with ASD from the non-ASD group. This novel system may provide a low-cost approach to early screening and identifying toddlers with ASD. However, machine learning is not as accurate as a human observer, and the detection of a single symptom like RTN is not sufficient enough to detect ASD.

## KEYWORDS

autism spectrum disorder (ASD), response to name (RTN), multimodal, machine learning, early screening, identification

## 1. Introduction

Autism spectrum disorder (ASD) is a set of heterogeneous neurodevelopmental conditions, characterized by early onset deficits in social communication, restricted interest, and repetitive behaviors (1). The prevalence of ASD has sustainably increased in recent decades. The latest epidemiological survey has reported that approximately one in 44 children (2.27%) was identified as having ASD in the United States (2). Previous studies have shown that the symptoms of ASD emerge at an early age (3), while the diagnostic age of ASD exists a significant delay. An investigation has demonstrated that the mean age at ASD diagnosis was 4.4 years, whereas the mean diagnostic delay was 2.2 years (4). Initiating early intervention can improve the prognosis of ASD (5), and the diagnostic delay implies the loss of the opportunity to receive early intervention for toddlers with ASD.

Response to name (RTN) is a crucial social response for toddlers. RTN plays an essential recognition functioning in social cognition and interaction. Toddlers aged 4–6 months listen significantly longer to their own names than others, suggesting RTN is an internalized behavior early in life (6). Toddlers with typical development (TD) significantly prefer their own names to others, especially those spoken by their mothers (7). Due to the crucial behavior in the development of social skills, the assessment of RTN may be useful in early screening and identification of ASD.

Many lines of evidence have shown a reduced RTN in toddlers with ASD (8, 9). A prospective study has examined the longitudinal patterns of RTN from 6 to 24 months and found that toddlers who consistently fail to respond to their names in the second year of life may be at risk not only for ASD but also for greater impairment in later age (10). The results suggest that persistent failure to RTN in early life may be a critical indicator for developmental disorders. Many screening and diagnostic instruments have included RTN as an important item (11, 12), such as the modified checklist for autism in toddlers (M-CHAT) and the autism diagnostic observation schedule (ADOS). Retrospective studies have found that diminished RTN relies merely on the parent's report or schedule (13, 14), whereas little is known to quantize RTN *via* the more accurate method.

The existing methods in screening ASD mainly rely on scale screening, such as M-CHAT and social communication questionnaire (SCQ) by parent's report, with some subjective errors and long time-consuming. Compared with the disadvantages of traditional methods, machine learning has risen to be a promising alternative in the screening and diagnosis of ASD (15, 16). Machine learning aims to construct predictive models from the datasets, which encompasses search methods, artificial intelligence, and mathematical modeling. Machine learning algorithms are applied as an intelligent method with minimal human involvement, with decision tree algorithms being used in data processing to detect ASD. A study has applied machine learning to distinguish ASD from TD children using a simple upper-limb reach-to-drop task (17), and the resulting model showed an accuracy rate of 96.7%, which suggest that machine learning could be a useful method of classification and discrimination in the diagnosis of ASD. Recently, Achenie et al. have demonstrated that the machine learning method was comparable to the M-CHAT with follow-up items in the accuracy of ASD diagnosis using fewer items (18). The results suggest that machine learning may be a promising tool in implementing automatic, efficient scoring in the diagnosis of ASD. A multimodal machine learning system

is a more advanced machine learning, which is a data-driven approach for early autism screening that combines general machine learning principles and behavioral analysis techniques to perform comprehensive computer analysis of multimodal audio–visual data. By extracting useful information and building complex models that surpass human performance in analyzing large datasets (19), MMLS can enhance our understanding of ASD and may further build a stronger foundation for early screening and identification.

To the best of our knowledge, toddlers with ASD are still mainly diagnosed by clinicians taking a long time to collect behavioral observations and case history. The early screening and identification of ASD are also major challenges to clinicians. Little is known about the question of whether MMLS could enhance or replace the role of clinicians in the screening and identification of ASD, and few studies have investigated the difference in RTN between ASD and non-ASD groups. The primary objective of our study was to explore the difference in RTN between toddlers with ASD and non-ASD groups using the MMLS. We aim to test the feasibility of the application of the machine approach in the rating works of RTN to predict ASD.

## 2. Materials and methods

### 2.1. Participants

We designed developmental delay (DD) and typical development (TD) as non-ASD groups. Toddlers with ASD and DD were randomly recruited from the inpatients or outpatients in Child Developmental and Behavioral Center, the Third Affiliated Hospital of Sun Yat-sen University in Guangzhou between April and November 2017. Meanwhile, we recruited toddlers with TD from the nearby community during the same period. Toddlers with ASD ( $n = 61$ ) were matched on chronological age to toddlers with DD ( $n = 31$ ) and TD ( $n = 33$ ), aged 16–30 months. The chronological age of children with ASD was 25.16 (3.71) months, and the sex ratio of men to women was 5.1:1. The chronological age of children with DD was 25.09 (3.68) months, and the sex ratio in this group was 5.2:1. The chronological age of children with TD was 24.73 (3.37) months, and the sex ratio was 5.6:1. There were no significant differences in chronological age and sex ratio among the three groups (all  $P$ -values > 0.05).

Toddlers with ASD were included to meet the diagnostic criteria of the diagnostic and statistical manual of mental disorders—the 5th edition (DSM-5) (20). Exclusionary criteria for toddlers with ASD included the following: (1) symptomatic autism, such as Rett syndrome, Fragile X syndrome, Angelman syndrome, and Prader–Willi syndrome, which were caused by known genetic defects or inherited metabolic diseases; (2) hearing impairment and moderate-to-severe sensory impairment; (3) cerebral palsy and poorly controlled seizures. The diagnosis of ASD in this study required a coincident diagnosis by two clinicians to ensure the quality and validity of the diagnosis.

Inclusion criteria for toddlers with DD included the following: (1) isolated developmental delay (involving single domain); (2) multiple developmental delays—two or more domains or developmental lines affected; (3) global developmental delay (GDD)—significant delay in most of the developmental domains (21). Exclusionary criteria for toddlers with DD included the following: (1) physical disabilities; (2) hearing impairment; (3) a history of serious brain diseases at birth, such as hypoxic ischemic encephalopathy and neonatal apoplexy.

Toddlers with TD were assessed by professional staff before the study, which did not have any confirmed or suspected developmental disorders in healthcare facilities. Exclusion criteria for toddlers with TD were as follows: (1) hearing impairment; (2) visual impairment; (3) speech and vocal impairment; (4) motor impairment due to any physical disability.

## 2.2. Procedures

To avoid habituation during frequent name calls, the evaluator or caregiver was requested to call the toddler's name no more than three times. The place of name call was chosen about 2 m behind the toddler according to the social distance (22). The evaluator first explained the experimental process and instructions to the caregiver and then guided the toddler going into the laboratory with a standard demonstration. The study began with toddlers playing with the preparative toy (jack in the box) placed on the table, with the caregiver sitting beside it. When the toddler's attention was completely attracted by the toy, the evaluator began to call his/her name behind. If the toddler did not respond to the initial name call, the evaluator paused for approximately 3 s and then called the toddler's name again. In this progress, up to three name calls were implemented until the toddler turned their head with eye contact or there was no response. The toddler was next allowed to freely play for 30 s. When the toddler's attention was focused on the toy again, the caregiver called the toddler's name by repeating the above-mentioned procedures, and the same is illustrated in [Figure 1](#).

## 2.3. Automatic name calling detection

An overview of the proposed multimodal "response to name" assessment framework is shown in [Figure 2](#). In our study, both speech processing and vision-based methods are incorporated to minimize human annotations in assessing responsiveness. In particular, we simultaneously consider response speed, response duration, and head and pose information to jointly determine a predicted score. A core task in "response to name" experiments is to locate the time stamps of name callings such that response latency can be measured. Name calling is annotated manually and incorporates unnecessary human interaction.

## 2.4. Data collection and processing

Behavioral data were obtained from eight high-resolution cameras which were evenly distributed around the laboratory. The voice was recorded by a digital audio recorder worn by the toddlers. The sensors are synchronized and are expected to comprehensively cover various modalities of the signals from the child. The assessment framework is illustrated in [Figure 2](#).

### 2.4.1. Clinical scoring

For the manual scoring method, the videos were played frame by frame through a free video player (the KM player). Every second of the video contains 30 frames. Items are scored from 0 to 2: "0" means the toddler had no response; "1" means the toddler turned their head toward the evaluator but without eye contact; "2" stands

for the toddler turned their head toward the evaluator and had eye contact with the evaluator.

Response time (RT) means the latency from the time which caller's voice was detected to the time which toddler's face was identified. If no response, we defined a maximal value of 10 s as RT. Response duration time (RDT) represents the time of video segments that contain the toddler's face. If no response, the RDT is 0. Only the response that appeared behind the name calling was viewed as valid. Furthermore, to compare the results of manual scoring and machine learning, two co-investigators and an expert in the diagnosis of ASD reviewed these videos, who were blind to the diagnosis and current group membership.

### 2.4.2. Automatic name calling detection

A core step of data processing was to locate the time stamp of the name calling. We designed an automatic speech recognition (ASR) system based on Kaldi, a toolbox widely used for speech recognition (23). In the registration step, each child's name was registered in the system. In addition, the ASR feature was matched to the system to obtain the time stamp during the name calling.

### 2.4.3. Face detection and recognition

Another critical step was to locate the child's face in the video. In this study, we used the DLib implementation of the face alignment methods proposed by Kazemi et al. to detect and align the faces simultaneously (24). Besides detecting faces, the algorithm returned 68 landmarks which later would be used to compute the head pose. Meanwhile, we also performed deep learning-based face recognition to identify a child's face. Similar to name calling detection, we registered each child's face in the system and verified every detected face based on the method.

### 2.4.4. Feature representation

With the detected child faces, we proposed a rule-based "response to name" assessment framework. Consisting with conventional clinical diagnosis, the framework was based on the following basic assumptions: The first assumption was that a clear response should happen with a relatively small latency upon calling. The second was that a clear response should last for a certain length of duration.

The response latency in the first assumption can be naturally modeled as follows:

$$\text{Latency} = T_f - T_{c1}$$

$T_f$  and  $T_{c1}$  indicate the time stamp of the first detected face and the beginning of the first call, respectively.

The response duration in the second assumption can also be modeled as follows:

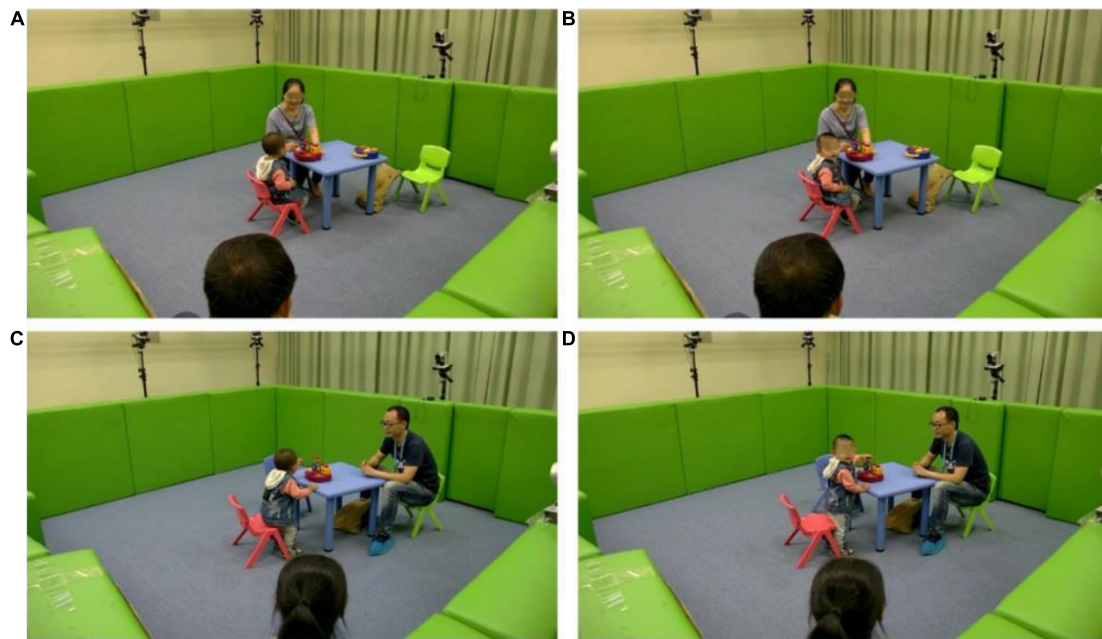
$$\text{Duration} = N_f / \text{Frame rate}$$

$N_f$  represents the total number of frames containing the detected child faces.

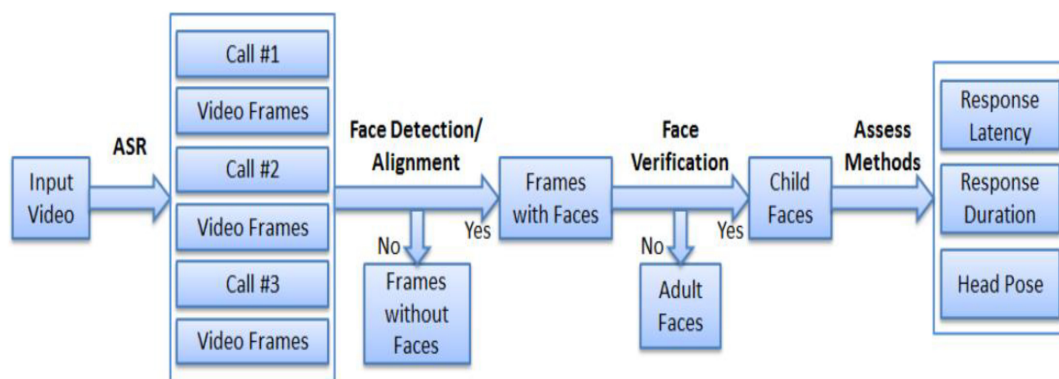
### 2.4.5. Score prediction

The data were divided into training and testing sets with a 10-fold cross-validation strategy, and then, we evaluated each subject in an iterative manner.

In this task, a positive response should have a relatively short response latency and a long duration of eye contact. In other words, a



**FIGURE 1**  
 Demonstration of different name call procedures. **(A)** Before the name call of the evaluator procedure. **(B)** After the name call of the evaluator procedure. **(C)** Before the name call of the caregiver procedure. **(D)** After the name call of the caregiver procedure.



**FIGURE 2**  
 Proposed multimodal machine learning framework toward RTN. \*ASR, automatic speech recognition; RTN, response to name.

typically developing test subject should respond to his name quickly and face the name caller for some time. We thus used response latency and front-facing duration as our features and applied the rule-based decision tree separately on each feature. Besides regressing to the clinical score, we also evaluated the performance of machine evaluation scores in direct ASD prediction. In our test protocol, the parent was first asked to participate in the name calling, followed by the doctor. Thus, there were two responses to name tests performed on each test subject. The ASD prediction result was jointly learned by these two tests.

## 2.5. Statistical analysis

Data management and analysis were performed by the statistical package for social sciences (version 20.0; SPSS Inc., Chicago, IL, USA). For quantitative data, normality tests were first performed.

Data that followed the normal distribution were statistically described by the mean and standard deviation (SD), and *t*-tests were applied to data comparison. Data that did not follow the normal distribution are described by the median (M) and interquartile range (IQR). Before group comparisons, the Kruskal–Wallis test was used to estimate the overall distribution. If the overall distributions were not identical, a multisample Kruskal–Wallis one-way ANOVA test by ranks was used to perform multiple comparisons between groups. Both quantitative and qualitative analyses were utilized to examine the current data. The quantitative analysis comprised independent samples *t*-tests and non-parametric tests. Qualitative analyses constituted the chi-square test. The difference in gender across groups was examined by the chi-square test. The difference in chronological age between the groups was examined using independent samples *t*-tests. Internal consistency reliability between two co-investigators was tested by the method of Cronbach. Receiver operating characteristic (ROC)



TABLE 1 Consistency of human and computer rating.

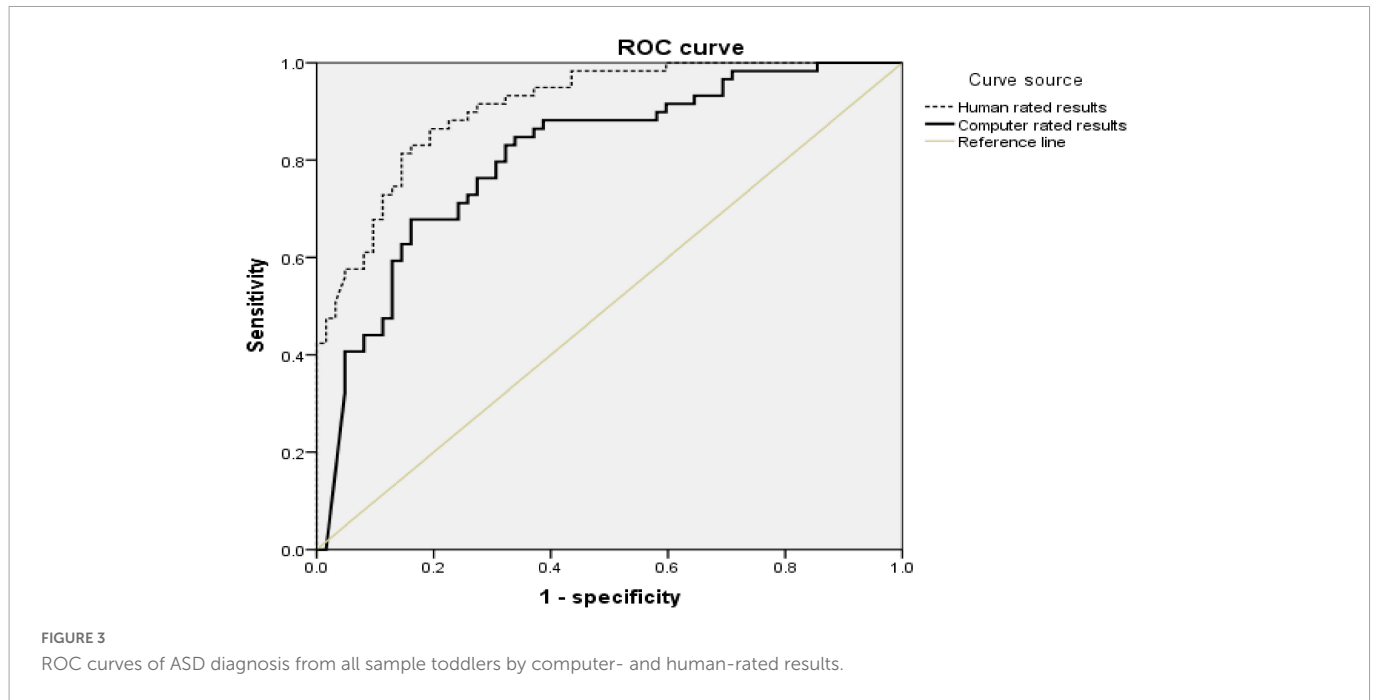
Item	Score	Evaluator procedure		Caregiver procedure	
		Human	Computer	Human	Computer
ASD		n = 61		n = 60	
	0	37 (60.66%)	36 (59.02%)	30 (50.00%)	27 (45.00%)
	1	2 (3.28%)	0 (0.00%)	3 (5.00%)	0 (0.00%)
	2	22 (36.07%)	22 (36.07%)	27 (45.00%)	25 (41.67%)
	Consistency	0.95		0.87	
DD		n = 31		n = 31	
	0	6 (19.36%)	6 (19.36%)	5 (16.13%)	4 (12.90%)
	1	2 (6.45%)	0 (0.00%)	1 (3.22%)	0 (0.00%)
	2	23 (74.19%)	21 (67.74%)	25 (80.65%)	24 (77.42%)
	Consistency	0.87		0.90	
TD		n = 33		n = 32	
	0	6 (18.18%)	6 (18.18%)	4 (12.50%)	4 (12.50%)
	1	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	2	27 (81.82%)	27 (81.82%)	28 (87.50%)	27 (84.38%)
	Consistency	1.00		0.97	
	Total consistency	0.92			

ASD, autism spectrum disorder; DD, developmental delay; TD, typical development.

TABLE 2 Computer-rated results of RTN among toddlers with ASD, DD, and TD.

Items	ASD	DD	TD	H value	P-value	Multiple comparison	Standard H value	Revised P-value
	M (IQR)	M (IQR)	M (IQR)					
<b>Evaluator procedure</b>								
Score	0	2	2	20.27	<0.001	ASD-TD	-4.12	<0.001
	(2)	(2)	(0)			ASD-DD	-3.10	0.006
						DD-TD	-0.75	1.000
Response time	10	3	2.17	25.98	<0.001	ASD-TD	4.51	<0.001
	(4.17)	(8.14)	(5.10)			ASD-DD	3.78	<0.001
						DD-TD	0.57	1.000
Duration time	0	1.88	1.73	26.53	<0.001	ASD-TD	-4.33	<0.001
	(1.22)	(2.31)	(2.52)			ASD-DD	-4.10	<0.001
						DD-TD	-0.13	1.000
<b>Caregiver procedure</b>								
Score	1	2	2	17.08	<0.001	ASD-TD	-3.35	0.002
	(2)	(0)	(0)			ASD-DD	-3.41	0.002
						DD-TD	0.15	1.000
Response time	9.05	2.00	1.80	19.14	<0.001	ASD-TD	3.69	0.001
	(7.31)	(3.59)	(6.82)			ASD-DD	3.46	0.002
						DD-TD	0.10	1.000
Duration time	0.10	1	1.45	10.90	0.004	ASD-TD	-2.84	0.013
	(1.48)	(2.04)	(1.52)			ASD-DD	-2.54	0.033
						DD-TD	-0.19	1.000

ASD, autism spectrum disorder; DD, developmental delay; TD, typical development. Revised P-value, P-value after Bonferroni correction; standard H-value, H-value after Bonferroni correction.



**TABLE 3** Area under the ROC curves of ASD diagnosis from toddlers with ASD, DD, and TD based on the human- and computer-rated RTN results.

Test variable	Area	Standard error	P-value	95% confidence interval	
				Lower limit	Upper limit
Computer rated results	0.81	0.03	<0.001	0.86	0.96
Human rated results	0.91	0.04	<0.001	0.73	0.89

analysis was used in predicting the diagnostic accuracy in toddlers with ASD. MedCalc 15.8 statistical software was applied to compare two ROC curves. In addition, accuracy, specificity, and sensitivity were conducted as an evaluation index in machine learning. A *P*-value of < 0.05 was considered statistically significant.

### 3. Results

#### 3.1. Consistency of RTN scores between the human rating and computer rating

Two colleagues rated the video data, respectively, and Cronbach’s alpha between colleagues was 0.98. Assessing results by human rating were used as the standard score. In the evaluator procedure, there were, respectively, 61 valid responses in ASD, 31 in DD, and 33 in TD. In the caregiver procedure, the valid responses were, respectively, 60 in ASD, 31 in DD, and 32 in TD. Taken together, the total consistency of the computer rating was 0.92 in these 248 valid responses as shown in [Table 1](#).

#### 3.2. Comparison of RTN indicators among ASD, DD, and TD

To study the response characteristics in different toddlers, response score, response time, and response duration time were evaluated. Both in the evaluator procedure and the caregiver

procedure, there were significant differences in the three indicators between ASD and TD, ASD, and DD (all *P*-values < 0.05), while the difference in the three indicators between DD and TD was not significant (all *P*-values > 0.05). The results are shown in [Table 2](#).

#### 3.4. ASD diagnostic efficiency, sensitivity, specificity, and accuracy based on the computer- and human-rated results of RTN

To evaluate the diagnostic efficiency of using computer rating results to detect ASD, ROC curve ([Figure 3](#)) was established for all data of RTN. We combined toddlers with DD and TD as a non-ASD group. Taking the results of both procedures into account, the area under the curve (AUC) was 0.81 ([Table 3](#)). The sensitivity, specificity, and accuracy of the ASD diagnosis based on the computer-rated results of RTN were, respectively, 80.0, 69.8, and 74.8% ([Table 4](#)). For human-rated results, the AUC was 0.91, and the sensitivity, specificity, and accuracy were, respectively, 83.3, 82.5, and 82.9% ([Table 5](#)).

MedCalc 15.8 statistical software was used to compare the AUC of different ROC curves. There was a significant difference between the AUC of the human-rated results and computer-rated results ( $Z = 2.71, P\text{-value} = 0.007$ ). The difference in accuracy between the computer diagnosis and human diagnosis was not significant ( $\chi^2 = 2.44, P\text{-value} = 0.118$ ).

**TABLE 4** Outcomes of ASD diagnosis based on the computer-rated results of RTN.

Item		Clinical diagnosis		Total
		ASD (n)	Non-ASD (n)	
Computer diagnosis	ASD	48	19	67
	Non-ASD	12	44	56
Total		60	63	123

ASD, autism spectrum disorder; Non-ASD, typical developmental and developmental delays.

**TABLE 5** Outcomes of ASD diagnosis based on the human-rated results of RTN.

Item		Clinical diagnosis		Total
		ASD (n)	Non-ASD (n)	
Human diagnosis	ASD	50	11	61
	Non-ASD	10	52	62
Total		60	63	123

ASD, autism spectrum disorder; Non-ASD, typical developmental and developmental delays.

## 4. Discussion

This study began with the aim of assessing the application of artificial intelligence in the early screening of children with ASD. The results indicate that the computer system can perform as effectively as humans in behavior coding. The present study investigated the early social skills of RTN among toddlers with ASD, TD, and TD. Compared with the non-ASD group, toddlers with ASD showed a reduced RTN call. Participants have rated automatically, which provides an advantage over previous screening methods to overcome the inaccuracy of manual scoring of video-based studies.

The total accuracy (92%) of the computer system was relatively high, which indicates that MMLS can accurately reflect the actual performance of different toddlers in RTN procedures. Among these toddlers, accuracy for toddlers with TD was the highest, especially in the evaluator's procedure. This phenomenon can be attributed to toddlers with TD being more easily adjusted to the new environment and more cooperative compared with the other two groups. Thus, their behavioral information was more easily and accurately captured. Nevertheless, toddlers with ASD had a tendency to perform hyperactively and exhibited more uncontrollable behaviors, which might make it more difficult to quantify behavioral data. Compared with the previous studies on the automatic detection of RTN, the current study used multimodal information to evaluate behavioral data, making the results more reliable.

Our results demonstrated that toddlers with ASD were more likely to fail to respond to name calls compared with the non-ASD group. A recent study has revealed that children with ASD responded to their names significantly less frequently than children without ASD using computer vision analysis (CVA). The CVA also exhibited that children with ASD who did orient to name call showed a longer latency before turning their head (25), which is consistent with our results. The diminished response can be explained by the theory of social motivation and social cognition in ASD. According to the social motivation theory, ASD was an extreme example of reduced social motivation (26). Decreased social orienting and social reward may be the main reason for the low response rate and long response latency to name call, and diminished social maintenance may be the main cause of shorter response duration time in toddlers

with ASD. Moreover, restricted patterns of interest in toddlers with ASD may further impact their interest in social stimulations, such as name call. Evidence has manifested that social cognition was positively associated with social functioning (27). Social cognition deficits may cause difficulties in toddlers with ASD to understand societal situations. As a result, toddlers with ASD did not realize that they should give a response to the name call and they did not know what kinds of response they should give, or they did not even understand that the name call was pointed to them. Other possible explanations for the poor response may be that toddlers with ASD have deficits in the ability of auditory spatial attention (28). Studies in adults with ASD also suggested that atypical auditory object processing had implications in understanding the communication difficulties of ASD (29).

Delays in identification and diagnosis have direct bad consequences in timely intervention for children with ASD. The later the initiating intervention, the less opportunity for toddlers with ASD to reach optimal outcomes. This machine learning provides a new approach to screening and identifying ASD, as it is able to identify toddlers with ASD at 2 years old. Furthermore, this machine learning is not difficult to carry out, without any harmfulness. The coding course is replaced by computers, reducing the difference in the evaluator's subjectivity. The present study showed that the difference in accuracy in the diagnosis of ASD between the computer and humans was not significant. The above-mentioned results suggest that computers and humans had similar performance in behavior coding. This finding has significant implications for developing an artificial intelligence test system in the early screening and diagnosing of ASD. The existing method in screening and prediction of ASD mainly relies on a scale screening and an empirical diagnosis by clinicians, which lacks objective tools. The current study provides a novel approach to coding video-based behavioral data objectively and intelligently, which may offer a complementary method for reliably and effectively assessing toddler behavior.

As a standardized paradigm, the present study has examined the feasibility of early screening of ASD using machine learning. In this approach, MMLS may save human resources in screening and diagnosing ASD. Moreover, it may provide a chance to be earlier identified for toddlers who have suffered a long waiting time before a comprehensive assessment in areas where medical resources are relatively deficient. There are still several limitations in our study. First, the current research is a cross-sectional study, while the development of social abilities in toddlers is a dynamic process. The social performance of one time point cannot reflect the whole societal ability. Second, our rating system could not completely discern all behavioral data as the experiment was conducted in a semi-structure environment and the toddlers must sit in a designated position, which might influence the behavioral performance of toddlers in a natural situation. Third, RTN cannot be used as a single indicator to predict ASD. The next step in future research ought to design a prospective method, including larger sample size and longitudinal follow-up. Moreover, machine learning is not as accurate as a human observer and the detection of a single symptom like RTN is not sufficient enough to detect ASD. The role of multiple social indicators of ASD should be involved in the future study. In addition, the procedure of RTN should be carried out under a free-play situation, where behaviors would be more natural and in line with ecological validity.

## 5. Conclusion

In summary, our results confirm that MMLS can accurately quantify behaviors in RTN procedures and effectively distinguish toddlers with ASD from the non-ASD group. This novel system may provide an accurate and low-cost approach to early screening and identifying toddlers with ASD.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by The Third Affiliated Hospital of Sun Yat-sen University. 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

F-LZ: conceptualization and methodology. S-HW: data curation and original—draft preparation. W-BL: visualization and investigation. H-LZ: software and supervision. ML: software and validation. X-BZ: reviewing and approval publication. All authors critically reviewed the manuscript drafts and approved the final manuscript.

## References

- Lord C, Elsabbagh M, Baird G, Veenstra-Vanderweele J. Autism spectrum disorder. *Lancet*. (2018) 392:508–20. doi: 10.1016/S0140-6736(18)31129-2
- Maenner MJ, Shaw KA, Bakian AV, Bilder DA, Durkin MS, Esler A, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveill Summ*. (2021) 70:1–16. doi: 10.15585/mmwr.ss7011a1
- Zablotsky B, Colpe LJ, Pringle BA, Kogan MD, Rice C, Blumberg SJ. Age of parental concern, diagnosis, and service initiation among children with autism spectrum disorder. *Am J Intellect Dev Disabil*. (2017) 122:49–61. doi: 10.1352/1944-7558-122.1.49
- Zuckerman K, Lindly OJ, Chavez AE. Timeliness of autism spectrum disorder diagnosis and use of services among U.S. Elementary school-aged children. *Psychiatr Serv*. (2017) 68:33–40. doi: 10.1176/appi.ps.201500549
- Dimian AF, Symons FJ, Wolff JJ. Delay to early intensive behavioral intervention and educational outcomes for a Medicaid-enrolled cohort of children with autism. *J Autism Dev Disord*. (2021) 51:1054–66. doi: 10.1007/s10803-020-04586-1
- Conine DE, Vollmer TR, Barlow MA, Grauerholz-Fisher E, Dela Rosa CM, Petronelli AK. Assessment and treatment of response to name for children with autism spectrum disorder: toward an efficient intervention model. *J Appl Behav Anal*. (2020) 53:2024–52. doi: 10.1002/jaba.737
- Imafuku M, Hakuno Y, Uchida-Ota M, Yamamoto J, Minagawa Y. “Mom called me!” Behavioral and prefrontal responses of infants to self-names spoken by their mothers. *Neuroimage*. (2014) 103:476–84. doi: 10.1016/j.neuroimage.2014.08.034
- Larsen K, Aasland A, Schjølberg S, Hansen UI, Diseth TH. Piloting the use of a short observation list for asd-symptoms in day-care: challenges and further possibilities. *J Autism Dev Disord*. (2020) 50:3413–23. doi: 10.1007/s10803-019-04313-5
- Nadig AS, Ozonoff S, Young GS, Rozga A, Sigman M, Rogers SJ. A prospective study of response to name in infants at risk for autism. *Arch Pediatr Adolesc Med*. (2007) 161:378–83. doi: 10.1001/archpedi.161.4.378
- Miller M, Iosif AM, Hill M, Young GS, Schwichtenberg AJ, Ozonoff S. Response to name in infants developing autism spectrum disorder: a prospective study. *J Pediatr*. (2017) 183:141–6.e1. doi: 10.1016/j.jpeds.2016.12.071
- Canal-Bedia R, Garcia-Primo P, Martin-Cilleros MV, Santos-Borbujo J, Guisuraga-Fernandez Z, Herraiz-Garcia L, et al. Modified checklist for autism in toddlers: cross-cultural adaptation and validation in Spain. *J Autism Dev Disord*. (2011) 41:1342–51. doi: 10.1007/s10803-010-1163-z
- Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule—generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Deve Disord*. (2000) 30:205–23. doi: 10.1023/a:1005592401947
- Osterling JA, Dawson G, Munson JA. Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Dev Psychopathol*. (2002) 14:239–51. doi: 10.1017/s0954579402002031
- Palomo R, Belinchon M, Ozonoff S. Autism and family home movies: a comprehensive review. *J Dev Behav Pediatr*. (2006) 27(2Suppl):S59–68. doi: 10.1097/00004703-200604002-00003
- Rahman MM, Usman OL, Muniyandi RC, Sahran S, Mohamed S, Razak RA. A review of machine learning methods of feature selection and classification for autism spectrum disorder. *Brain Sci*. (2020) 10:949. doi: 10.3390/brainsci10120949
- Aglinskas A, Hartshorne JK, Anzellotti S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science*. (2022) 376:1070–4. doi: 10.1126/science.abm2461

## Funding

This study was supported by the National Natural Science Foundation of China (No: 81873801), Science and Technology Program of Guangzhou, China (No: 202007030011), and the Guangdong Medical Research Foundation, China (No: A2022039).

## Acknowledgments

We are grateful to all colleagues for their assistance in participant recruitment and test arrangement. We sincerely thank the caregivers and toddlers who participated in this study.

## Conflict of interest

The 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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



17. Crippa A, Salvatore C, Perego P, Forti S, Nobile M, Molteni M, et al. Use of machine learning to identify children with autism and their motor abnormalities. *J Autism Dev Disord.* (2015) 45:2146–56. doi: 10.1007/s10803-015-2379-8
18. Achenie LEK, Scarpa A, Factor RS, Wang T, Robins DL, McCrickard DS. A machine learning strategy for autism screening in toddlers. *J Dev Behav Pediatr.* (2019) 40:369–76. doi: 10.1097/DBP.0000000000000668
19. Liu W, Zhou T, Zhang C, Zou X, Li M. Response to name: a dataset and a multimodal machine learning framework towards autism study. *Proceedings of the 7th International Conference on Affective Computing and Intelligent Interaction.* San Antonio, TX (2017).
20. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5.* 5th ed. Washington, DC: American Psychiatric Association (2013). doi: 10.1176/appi.books.9780890425596
21. Khan I, Leventhal BL. *Developmental Delay StatPearls.* Treasure Island, FL: StatPearls Publishing (2021).
22. Hall ET. The silent language. *Anchor Books.* (1980) 38:87–96.
23. Povey D, Ghoshal A, Boulianne G, Luká B, Vesel K. The Kaldi speech recognition toolkit. *Proceedings of the IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.* Waikoloa Village, HI (2011).
24. Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition.* Columbus, OH (2014). doi: 10.13140/2.1.1212.2243
25. Perochon S, Di Martino M, Aiello R, Baker J, Carpenter K, Chang Z, et al. A scalable computational approach to assessing response to name in toddlers with autism. *J Child Psychol Psychiatry.* (2021) 62:1120–31. doi: 10.1111/jcpp.13381
26. Chevallier C, Kohls G, Troiani V, Brodtkin ES, Schultz RT. The social motivation theory of autism. *Trends Cogn Sci.* (2012) 16:231–9. doi: 10.1016/j.tics.2012.02.007
27. Bishop-Fitzpatrick L, Mazefsky CA, Eack SM, Minshew NJ. Correlates of social functioning in autism spectrum disorder: the role of social cognition. *Res Autism Spectr Disord.* (2017) 35:25–34. doi: 10.1016/j.rasd.2016.11.013
28. Soskey LN, Allen PD, Bennetto L. Auditory spatial attention to speech and complex non-speech sounds in children with autism spectrum disorder. *Autism Res.* (2017) 10:1405–16. doi: 10.1002/aur.1790
29. Lodhia V, Hautus MJ, Johnson BW, Brock J. Atypical brain responses to auditory spatial cues in adults with autism spectrum disorder. *Eur J Neurosci.* (2017) 47:682–9. doi: 10.1111/ejn.13694