scientific reports

OPEN

Risk assessment and automatic identification of autistic children based on appearance

Ruisheng Ran[1], Wei Liang[1], Shan Deng[1], Xin Fan[2]✉, Kai Shi[1], Ting Wang[1], Shuhong Dong[1], Qianwei Hu[1] & Chenyi Liu[1]

The diagnosis of Autism Spectrum Disorder (ASD) is mainly based on some diagnostic scales and evaluations by professional doctors, which may have limitations such as subjectivity, time, and cost. This research introduces a novel assessment and auto-identification approach for autistic children based on the appearance of children, which is a relatively objective, fast, and cost-effective approach. Initially, a custom social interaction scenario was developed, followed by a facial data set (ACFD) that contained 187 children, including 92 ASD and 95 children typically developing (TD). Using computer vision techniques, some appearance features of children including facial appearing time, eye concentration analysis, response time to name calls, and emotional expression ability were extracted. Subsequently, these features were combined and machine learning methods were used for the classification of children. Notably, the Bayes classifier achieved a remarkable accuracy of 94.1%. The experimental results show that the extracted visual appearance features can reflect the typical symptoms of children, and the automatic recognition method can provide an auxiliary diagnosis or data support for doctors.

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that typically appears in early childhood development. The disorder is characterized by three main symptoms: social impairment, speech impairment, and behavioral abnormalities. Social impairments are central to autism[1,2]. Autism is diagnosed primarily by professional doctors and diagnostic scales. The Diagnostic Checklist for Children with Behavioral Disorders (DCBDC) was the first attempt to assess children with autism[3]. After that, many diagnostic assessment scales were successively proposed, including CABS[4], ABC[5], ADOS[6], and CHAT[7] etc. These traditional methods may have limitations such as subjectivity, time and cost[8].

In an early study, Zwaigenbaum et al. analyzed home videos of children with autism during the undiagnosed period. The results showed that children with autism and typically developing (TD) children could be distinguished by specific facial features, such as gaze concentration, call response, and facial expressions[9]. With the development of artificial intelligence technology, computer vision and machine learning have become new and effective methods to assess and diagnose autism[10]. In recent years, computer vision-based assessments of children with autism have mainly included:

1. Analysis based on facial expressions or emotion. Manfredonia et al.[11] used facial expression analysis software (FACET) to study the ability to facial expression in autistic patients; Egger et al.[12] developed a mobile app to collect children's videos and analyze children's emotions and behaviors with computer vision methods; Paolucci et al.[13] used three core sections of ASD—emotional, sensorimotor, and behavioral and then to make an early prediction.

2. Analysis based on eye gaze and attention. Pierce et al.[14] studied autistic children's attention by tracking eye movements with an eye tracker; Bovery[15] collected gaze and attention data using convenient devices such as cameras or smartphones and used computer vision algorithms to analyze the attention of autistic children; Vabalas et al.[16] studied the similarity of imitation of motion between autistic and normal adults, and classified normal adults and autistic patients using eye gaze features. Alvari et al.[17] designed a software tool "EYE-C"; Based on eye-tracking data, Wei et al. used machine learning models such as random forests to detect ASD[18].

3. Analysis based on head pose. Martin et al.[19]used the computer vision method to analyze attention by the head movements of the children in social interaction. Dawson et al.[20] used computer vision to assess midline head posture, then quantified changes in early childhood motor behavior, and finally provided a new automated method for autism risk identification. Zhao et al.[21] used head movement features to identify ASD.

[1]The College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China. [2]The Department of Child Health Care, Chongqing Health Center for Women and Children, Chongqing 400010, China. ✉email: 2916497815@qq.com

4. Analysis based on children's speech or phonology. Because prosodic abnormalities or spectrum disorders often occur in the speech intonation of autistic patients, Yasushi et al.[22] used machine learning methods to compare the speech of children with ASD and TD. Ramesh et al.[23] used the voice data of autistic children and typical development children from TalkBank, the largest database of spoken languages, to develop machine learning models to predict ASD.

In addition, some studies integrate facial features into the diagnosis of autism. Hashemi et al.[24] collected behavioral features of children such as head posture and gaze with cameras, used computer vision methods to perform the quantitative calculation of the above features, and then presented the assisted diagnosis of autism. Jiang et al.[25] proposed a machine learning classification model based on gaze features and facial emotion recognition. Jaiswal et al.[26] used a depth measurement camera (Microsoft Kinect) to collect videos of subjects listening to stories and answering questions, used expression analysis and 3D behavior analysis to extract facial features, and finally used a machine learning method for the auxiliary diagnosis of autism. Li et al.[27] classify children with ASD in video sequences by integrating appearance-based features of facial expressions, head poses, and head trajectories.

Although some studies have used artificial intelligence to evaluate and analyze children with autism, these methods have some defects. Some methods require extensive instruments, such as eye trackers and Magnetic Resonance Imaging (MRI). Other methods only use certain features of children, such as gaze or head posture, which may not provide a complete picture of the child's condition. Therefore, more research is needed to improve the analysis, modeling, and classification of children with autism.

In this paper, based on the appearance features of children, we used computer vision methods to evaluate children with autism and then identify them. First, we designed an experimental material that contains six segments. Then, we recorded the facial video of the children while they watched the experimental material and established an Autism Children's Faces Dataset (ACFD). Social impairments are central to autism, they may appear dull in emotional expression, or even have no obvious emotional expression and often show atypical attention and gaze. They are uninterested, dull and indifferent to name calls. As the central nervous system may be abnormal, autistic children have less head movement[2,13,21,25,28]. Based on this, we extracted some appearance features of the children: facial appearing time, concentration analysis from eye gaze, name calls response time, and emotion expression ability based on the computer vision methods. Finally, we combined these features and used machine learning methods to identify children with autism.

Results

Statistical analysis

Difference analysis

In the Feature Modelling Section, we evaluate the children from the appearance of children: facial appearing time, concentration analysis from eye gaze, name calls response time, and emotion expression ability, and the corresponding six parameters R_{face}, G_{steady}, G_{follow}, T_{react1}, T_{react2}, S_{happy}. In that section, we have modeled the assessment metrics for children with ASD and given some examples. But these examples are only comparisons between an individual autistic child and a normal child. In this section, we make a difference analysis for six parameters on the dataset ACFD. The results are shown in Fig. 1 and Table 1.

The mean standard deviation and *t*-test results of six parameters for the two groups are shown in Table 1.

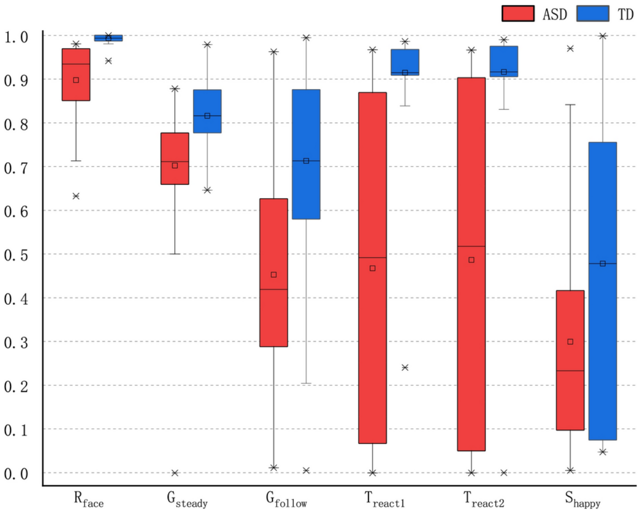


Figure 1. The box plot of the difference analysis of six feature parameters in the ACFD dataset. The horizontal axis represents six feature parameters, and the vertical axis represents the interval [0,1]. In which the upper and lower sides of the box are quartile spacing, the horizontal line is the median, the rectangle is the mean, the extension line is the upper edge (Q3+1.5IQR), and the lower edge (Q1-1.5IQR), "*" represents the outliers in the data.

Feature	ASD, M (SD)	TD, M (SD)	*t*-test
R_{face}	0.92 (0.09)	0.99 (0.01)	− 8.33***
G_{steady}	0.46 (0.24)	0.71 (0.22)	− 7.47***
G_{follow}	0.72 (0.13)	0.82 (0.07)	− 6.78***
T_{react1}	0.48 (0.38)	0.92 (0.11)	− 10.49 ***
T_{react2}	0.49 (0.39)	0.92 (0.15)	− 9.60***
S_{happy}	0.31 (0.25)	0.48 (0.35)	− 4.02***

Table 1. The statistical analysis of six parameters. Note: M is an abbreviation of "Mean value", and SD is the abbreviation of "Standard deviation". In the *t*-test, the *t* value reflects whether there is a difference between the two variables, and the sign reflects the direction of the difference. The *p*-value reflects the significance of the difference. The smaller the *p*-value is, the more significant the difference is. "*" represents $p \leq 0.05$,"**" represents $p \leq 0.01$, and "***" represents $p \leq 0.001$

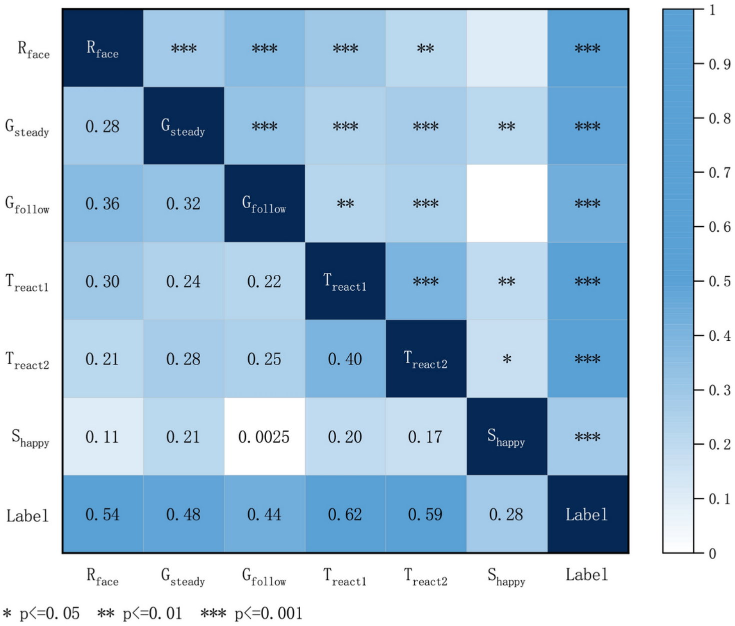


Figure 2. The matrix thermal diagram of Pearson correlation coefficient and *t*-test, in which the lower left part is the correlation coefficient *r*, and the upper right part is the significance mark of *p*-value in *t*-test between the six features, or between each feature and label.

As seen in Fig. 1 and Table 1, we can see the statistical regularities of the data distribution: for the six parameters, the mean values of children with ASD are all lower than those of children with TD, and the variance values of children with ASD are greater than those of children with TD. In the *t*-test, the *t* values are all negative, showing that the data distribution directions of children with ASD and children with TD are opposite; and the *p* values are all small, showing that there is a significant difference between the two. This shows that: (1) In the above all aspects, the children with ASD show much worse than the children with TD. (2) It is reasonable to use the above six parameters to evaluate children with ASD, which helps distinguish ASD children from TD children.

Correlation analysis

In this section, the correlation analyses of six feature parameters are made to find whether there is a correlation between these features or between features and tags.

We first mark autism as 0 and non-autism as 1. We compute the Pearson correlation coefficient of the six features, between each feature and label. Then we show the matrix thermal diagram of Pearson correlation coefficient and *t*-test in Fig. 2, in which the lower left part is the correlation coefficient *r*, and the upper right part is the significance mark of *p*-value in the *t*-test of the six features, or between each feature and label.

From Fig. 2, except that there is a correlation between R_{face} and G_{steady} (r = 0.42, p = < 0.001), the correlation between each feature is low and there is no linear correlation. It shows that the six features are independent of each other. From the bottom row and the last column of Fig. 2, among the six features, G_{follow}

, $T_{Treact1}$, $T_{Treact2}$, and S_{happy} are significantly correlated with the autism label. It shows that the six features can distinguish the children with ASD from the children with TD.

Classification results

On the ACFD dataset, the six features are extracted by the method in the Feature Analysis Section. The six features are concatenated as a feature vector and then it is used as the child's features. To verify the validity of the features, in this section, we use different classifiers to classify ASD and TD based on these features for diagnosis. Where, the Support vector machine (SVM), *k*-Nearest Neighbor classifier (KNN), Decision Tree (DTree), Linear Regression classifier (LR), Bayesian classifier (Bayes), Random Forest (RF), and Linear Discriminant Analysis (LDA) are used to classify, and the Leave-One-Out Cross-Validation is used to evaluate the classification performance of the model.

As seen from the "Video collection of children" Section, we have recruited 92 children with autism and 95 TD children. For this issue, the Leave-One-Out Cross-Validation method is to use one participant as a test set on the entire ACFD set, and all the remaining participants as a training set, then select the next participant as a test set and the remaining participants as a training set, and so on. Then, the average of all the tests is taken as the final result.

We then group the participants into four categories according to their ground truth (actual positive or negative) and the predicted group membership (predicted positive or negative), which are summarized as the following confusion matrix as shown in Table 2. Where positive refers to the ASD group, while negative refers to the TD group.

Where, the accuracy, recall rate, and F1-score are used to evaluate the classification model. The precision is calculated as the count of the correctly predicted participants divided by the count of all participants, i.e.,

$$Accuracy = \frac{TP + TN}{N}$$

Sensitivity (also known as True Positive Rate, TPR or Recall) is the proportion of samples that are actually positive cases (ASD) and predicted to be positive cases (ASD), measuring the model's ability to correctly identify positive cases (ASD), i.e.,

$$Sensitivity(R) = \frac{TP}{TP + FN}$$

Specificity is the proportion of samples that are actually negative cases (TD) and predicted to be negative cases (TD), measuring the model's ability to correctly identify negative cases (TD), i.e.,

$$Specificity = \frac{TN}{TN + FP}$$

Positive predictive value (PPV, also known as Precision) is the proportion of samples predicted as positive cases (ASD) that are truly positive cases (ASD). The higher the precision, the more accurate the model's predictions are, i.e.,

$$PPV(P) = \frac{TP}{TP + FP}$$

Negative Predictive Value (NPV) refers to the proportion of samples predicted as negative cases (TD) that are actually negative cases (ASD). It reflects the model's ability to correctly exclude TD, i.e.,

$$NPV = \frac{TN}{TN + FN}$$

The F1-score is an indicator of the harmonic average of precision and recall. The higher the F1-score, the better the classification performance of the model, i.e.,

$$F1 = \frac{2 \times P \times R}{P + R}$$

	Predicted membership	
Ground truth	**Positive (ASD)**	**Negative (TD)**
Positive (ASD)	*TP*	*FN*
Negative (TD)	*FP*	*TN*

Table 2. The confusion matrix. The confusion matrix is used to display the number of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) classification results in classification models.

Model	Accuracy	Sensitivity (R)	Specificity	PPV (P)	NPV	F1-score
SVM	90.9	94.8	86.5	88.5	93.9	91.5
KNN	88.7	93.8	83.1	85.8	92.5	89.7
DTree	88.2	89.7	86.5	87.9	88.5	88.8
Bayes	94.1	94.8	93.3	93.9	94.3	94.4
LR	89.8	94.8	84.3	86.8	93.8	90.6
RF	93.5	91.8	95.5	95.7	91.4	93.7
LDA	91.9	96.9	86.5	88.7	96.3	92.6

Table 3. The performance of model (%).

Model	Accuracy	Sensitivity	Specificity	Feature	Classifier
Jiang et al. (2019)[25]	86.2	91.3	–	Gaze and emotion	
Vabalas et al. (2020)[16]	78.0	57.0	–	Kinematic and eye movement	
Ali et al. (2022)[29]	86.4	–	–	Behavior	Multi-modality fusion network based on 3D-CNN
Zhao et al. (2022)[21]	92.1	88.9	95.0	Head movement	DT
Varghese et al. (2023)[30]	90.0	–	–	Head pose and motion parameters	AutoEncoder
Paolucci el al. (2023)[13]	–	89.0	86.0	Sensorimotor, behavioural, emotional	XGBoost
Wei et al. (2024)[18]	76.9	83.1	69.4	Eye-tracking data	Random forest
Ours	94.1	94.8	95.3	Face-based feature	Bayes

Table 4. The comparison with some latest studies (%).

From Table 3, when a Bayes classifier is used the performance of the model is best, i.e., the accuracy can reach 94.1%, the recall is 94.8%, the PPV is 93.9% and the F1-score is 94.4% of the classification model.

We also compare our method with similar studies in recent years. These previous studies were mainly based on head movement, emotional expression, behavior patterns, etc., using a variety of machines learning models, including decision tree, xgboost and deep neural network. The comparison results are shown in Table 4 . It can be seen that the proposed method has advantages, which proves its effectiveness and potential applications for further development.

Discussion

This study has three contributions: (1) the construction of a dataset of faces of autistic children (ACFD). We design interactive scenes using representative video segments, storylines, and calling children's names. And then we collect about 7 minutes of videos for each child, which captures appearance information such as children's facial expressions, eye gaze, and head posture. Thus, an Autism Children's Faces Dataset (ACFD) is constructed, which contains 187 children (92 autistic children and 95 typically developing children). (2) Risk assessment of children with ASD. With the computer vision method, we get appearance features of children from four aspects: facial appearing time, concentration analysis from eye gaze, name calls response time, and emotion expression ability. These features can be used as measurable indicators of children with ASD and can effectively express the potential relationship between autistic children and symptoms. (3) The automatic identification of autistic children based on machine learning. We combine the features mentioned above and use machine learning models to distinguish children with autism from children with TD. The accuracy of the classification model can reach 94.1% (when the Bayes classifier is used). This novel method is a relatively objective, fast and cost-effective approach compared to traditional diagnosis methods.

In the future, this research can be studied from the following aspects: (1) More comprehensive evaluation may be used. For example, for the concentration analysis, except the yaw angle of the gaze in this study, the pitch angle can also be considered; In terms of facial emotion analysis, except the happy emotions in this research, boredom and disgust, etc. expressions can also be considered in the follow-up research. (2) The classification results with respect to the automatic identification of children. Since the dataset cannot be directly published, it is difficult to make a direct comparison. Researchers interested in this dataset can access it, provided that they have signed a confidentiality agreement prior to use. In the future, we will also use computer vision technology to develop a numerical feature dataset of children with ASD, and publish it for scientific research. (3) Autism tends to be very difficult to diagnose because it is often mistakable with other developmental conditions such as ADHD, SLP, etc. So, a third group who are negative for ASD yet not TD may be considered. (4) Due to the complexity inherent in autism, the data set is acknowledged to be not representative of the whole autism. To overcome this challenge, some novel technologies, for example zero-shot learning, may be as a potential solution.

Methods

Statement: this study underwent prospective review and approval by the Medical Ethics Committee of Chongqing Health Center for Women and Children. Informed consent was obtained from all participants and/or their legal guardians. All methods were performed in accordance with the relevant guidelines and regulations.

Data collection

Experimental material

This research aims to extract typical features of autistic children from the appearance such as children's facial expressions, gaze, etc., and then use computer vision methods for autism risk assessment, followed by automatic identification using machine learning methods. The appearance features include facial appearing time, concentration analysis from eye gaze, name calls response time, and emotion expression ability. To obtain these, it is necessary to first design a social interaction scene. We created a 7-min and 23-s video as the experimental material with representative video segments, storylines, and calling the child's name as the interactive scene. The children are then asked to view the video in front of the laptop screen, while their facial information is recorded using the camera on the laptop.

The experimental materials designed and produced in this paper include the following six segments:

1. The first segment is a 30-s video of the balloon rising from five evenly spaced positions at the bot-tom of the screen to the top.
2. The second segment involves a name-call test that lasts 10 s. After being prompted by the screen, the experimenter calls the child's name from the side, and the child responds by turning their head towards the call. The experimenter records the time when the child responds to the call.
3. The third segment is a video of a ball bouncing back and forth for 30 seconds. The ball moves from the left side of the screen to the right side and back to the left side, back and forth twice.
4. The fourth segment is the second name-call test, which lasts 10 s. The content and operation of this segment are the same as that of the second segment.
5. The fifth segment is a one-minute video showing the movement of the left and the silence of the right.
6. The sixth segment consists of two animations with a duration of 5 min. These animations are selected under the guidance of professional doctors to ensure that the content is engaging and evokes positive emotions in children.The first to fifth segments of the experimental materials are designed and produced by our research group. The production of the first, third, and fifth video segments required balloons, balls, and toys, respectively. After purchasing these materials, we create videos. The second and fourth video segments involve a name calls test, in which a person in our research group calls the child's name on the spot. The sixth segment is collected from the Internet and is open access.

Video collection of children

We collect videos of children with autism at Chongqing Health Center for Women and Children, China. Before collection, this study underwent prospective review and approval by the Medical Ethics Committee of Chongqing Health Center for Women and Children. When collecting, we will inform parents/guardians of the experimental purpose and content, and sign a written informed consent form with them after obtaining their consent. Each child is assessed firstly with the Gesell Developmental Schedules[31]. Then, according to the results of the scale and clinical experience, if the diagnosis of ASD is suspected, the children are further assessed with the Modified Checklist for Autism in Toddlers (MCHAT)[32] or the Checklist for Autism in Toddlers 23 (CHAT-23)[33]. Finally, a final diagnosis is assessed with the Autism Diagnostic Observation Schedule (ADOS)[34] and the child's development and other diagnostic tests. In this research, videos of 92 children diagnosed with autism, including 68 males (73.9%) and 24 females (26.1%) children, are collected. As seen, the proportion of male children in the ASD group is higher. Some study has shown that the incidence of ASD in boys is four times higher than that in girls1[1], which has also been confirmed in this ASD group. The ASD group are all pre-school children, aged between 20 and 60 months, with an average age of 33.4 months for males and 31.5 months for females.

We also collected videos of 95 children with TD to compare with ASD. These children are recruited from a kindergarten in Chongqing, China. Before collecting, we also sign the written informed consent form with the parents/guardians after getting their consent. The children are observed and diagnosed by doctors and none of the children have language barriers, behavioral disorders, or other autism symptoms. The age of the TD group is between 24 and 60 months, similar to that of ASD group.

The data collection process involves using a laptop equipped with a 1080p high-resolution camera and microphone, a desk, and a chair. A quiet and comfortable environment is selected for the experiment. During data collection, the laptop is placed on the desk and the chair is placed approximately 50 cm in front of the computer screen. The child is then asked to sit alone or accompanied by their parents in front of the laptop screen. The child is shown the experimental material while their facial sensory information is recorded using the camera on the laptop. When the child finished watching, the video recording for that child ended. The experimenters always stand on the left and right sides of the child.

Autism children face dataset

The facial videos of children are processed by filtering out unfinished experiments and video content with a non-standard collection process, masking and coding the faces of the parents in the videos, and removing the noisy video during the name calls test. A table is used to record the child's age, sex, time of two name-calls, diagnosis results from a doctor, and other necessary information. Finally, an Autistic Children Face Dataset (ACFD) is built, which contains 187 (corresponding to 187 children) of about 7.5 minutes of children's facial video and a data presentation table. ACFD is described in Table 5.

The children's dataset ACFD will be securely stored on a computer server in our laboratory for long-term research work. Due to the personal privacy of children, we promise not to disclose their identity information and videos to anyone outside the research team. Researchers interested in this dataset can access this dataset, provided that they have signed a confidentiality agreement prior to use. Regarding the rules for obtaining and using the dataset, please refer to section "Data availability".

Facial feature extraction and modelling

Due to the social disorder of autistic children, they may appear dull in emotional expression, or even have no obvious emotional expression, and they may show less facial expression changes such as smiling and frowning. Autistic children often show atypical attention and gaze, and are difficult to maintain attention to a task for a long time, and are vulnerable to external influence, or have no clear attention goal. Due to social barriers or limited communication ability, autistic children may be slow, indifferent or lack of interest in name calling. As the central nervous system may be abnormal, the facial muscle control of autistic children may be impaired, resulting in reduced head movements[2,13,21,25,28]. Based on this, this research evaluates the ASD children and the TD children in four aspects: facial appearing time, eye concentration analysis, the response time for name calls and facial emotion expression.

Since the facial data recorded in ACFD is original biological information, in this study, the facial features of children are extracted by computer vision methods, then they are modeled to obtain the visual features of children, which are then used for evaluation and identification of children. The feature extraction and modeling are described below.

Feature extraction

In this study, four computer vision techniques[35]: facial landmark, gaze estimation, head pose estimation and facial action unit analysis (AU) is used to extract virtual features of children.

- Facial landmark detection

  In this study, the Convolutional Experts Constrained Local Model (CE-CLM)[36] is used to calculate the location distribution of facial landmarks. Because the facial organs are stable, the position of landmarks in facial structure is relatively constrained. By evaluating the alignment probability of mark points in each pixel, each landmark can be accurately located.

- Head pose estimation

  The facial landmark extracted from the CE-CLM model is represented by three-dimensional coordinates. To solve the n-point problem in perspective[37], the facial landmark positions are projected into a two-dimensional coordinate system using an orthographic camera projection. The vector coordinates and deflection angle of the head pose in 3D space are obtained by taking the head pose in front as the reference vector. The directional angle of the head pose includes the Yaw, Pitch, and Roll angle.

- Gaze estimation

  A 3D coordinate system with the camera center as the origin is first established, then a 3D eye model is established by landmarks of the eye, and the center coordinates of the eye sphere and the center coordinates of the pupil are calculated. The 3D vector from the center of the eye to the center of the pupil was the eye gaze vector[38]. Then, by the coordinate point position of the eye center, the variation of the eye center, the variation of yaw and pitch angles of the direction vector is calculated and then the gaze direction is gotten.

- Facial action unit recognition

  Facial Action Unit, referred to as AU, is a facial expression label defined by Facial Action Coding System (FACS)[39]. A combination of one or more facial action units can represent a person's facial emotions. Baltrušaitis et al. divide the facial image into several sub-regions and align the image according to the preset size,

Information	Gender	Group	
		ASD	TD
Participants (%)	Male	68 (73.9%)	54 (56.8%)
	Female	24 (26.1%)	41 (43.2%)
Average age month (SD)	Male	33.4 (12.73)	46.1 (17.28)
	Female	31.5 (8.21)	48.6 (16.93)

Table 5. Description of ACFD.

extract the facial Histogram of Oriented Gradient(HOG) feature, use the principal component analysis (PCA) method to extract the dimension of the HOG feature vector, and finally get the AU unit[40].

Feature modelling

From the facial features extracted in the previous section, this section makes modeling from four aspects: facial appearing time, concentration analysis from eye gaze, name calls response time, and emotion expression ability. Based on this, we conduct autism risk assessment and automatic identification for children.

- Define

  This study has the following definitions:

  1. The first to sixth video segments of the experiment material is denoted as F_1-F_6, respectively. $card(F_i)$ is the number of frames in the video segment F_i.
  2. f_i represents the *i*th frame image.
  3. $conf(f_i)$ represents the confidence level of face detection for the *i*th frame image.
  4. $gaze\psi(f_i)$ represents the yaw angle of the gaze vector tin the horizontal direction for the *i*th frame image, and $gaze\psi(f_i)'$ represents the derivative of the yaw angle of the *i*th frame.
  5. $pose\psi(f_i)$ represents the yaw angle of the head pose in the horizontal direction for the *i*th frame image.
  6. $AU_j(f_i)$ represents the *j*th facial action unit value of the *i*th frame image.
  7. The normalized sigmoid function is formulated as: $y = 1/(1 + \exp(-w(x - t)))$, where $w = 1$, and the parameter *t* represents the threshold value of the sigmoid function. If *x* is mapped into the interval [0.5, 1], and $x < t$, *x* is mapped into the interval [0, 0.5], therefore, the threshold *t* is taken as the cut-off point, and the function has the property of binarization.

- Facial appearing time

  When subjects look at experimental materials in front of the computer, generally speaking, TD children can stay always in front of the computer, but some ASD children may not participate throughout the experiment due to lack of concentration. The facial appearing time is the duration of time that the subject's face fully appears on the computer screen while watching the experimental material. This feature indicates whether the subject is constantly watching the experimental materials, thus reflecting the level of concentration and engagement of the children.We use the face detection algorithm in computer vision to detect the faces of the children in each frame of the video. The face detection reliability of each frame of the video is calculated as $conf(f_i)$, which takes in the interval [0,1]. The threshold $\varepsilon = 0.77$ was set by the comparison experiment with manual annotation. If the reliability $conf\ (f_i) \geq \varepsilon$, it was considered that the current frame has a face; otherwise, it was considered that there is no face. So, the facial appearing time can be formulated as:

$$R_{face} \ = \ \frac{card\left(\{conf\ (f_i) \geq \varepsilon, f_i \in F\}\right)}{card\left(F\right)} \tag{1}$$

R_{face} represents the ratio of the number of frames where faces appear to the total number of frames. The larger the R_{face} value is, the longer the facial appearing time is and the higher the child's involvement is. The smaller the R_{face} value is, the lower the child's involvement is.

In the experiment, an ASD child and a TD child are randomly selected to analyze their face detection reliability and R_{face} value, as shown in Fig. 3. The R_{face} of the ASD child is 0.75, and that of the TD child is 0.99, showing that the involvement of the ASD group is much worse than that of the TD group.

- Concentration analysis from eye gaze

  In this section, we analyze children's concentration based on the yaw angle of the eye gaze. The yaw angle reflects the subject's gaze angle in the horizontal direction. We evaluated the gaze stabilization ability with the video of the children watching the segment F_1, and evaluated the gaze tracking ability with the video of the children watching the segment F_3 .
  (1).Gaze stabilization ability
  The segment F_1 is a video of children observing the balloons rising. When a subject watches the segment F_1 , his gaze should follow the balloon up and not sway horizontally, then his yaw angle in the horizontal direction should be stable. So, the yaw angle can be used to analyze the gaze stabilization ability of the subject. The smaller the yaw angle, the stronger the gaze stability. The illustration of the segment F_1 and the yaw angle of the subject is shown in the following Fig. 4.

There are 5 upward movements of the balloon on the segment F_1 . For each rise of the balloon, the yaw angle of gaze is calculated and then there are five groups of the yaw angle data, as shown in Fig. 5. In this figure, for each group, the yaw angle curve in ASD (red line) is more volatile than that in TD (blue line). It shows that when the ball goes up, the yaw angle of the ASD child sway from side to side, and the gaze stabilization ability of the ASD child is weaker than that of the TD child.

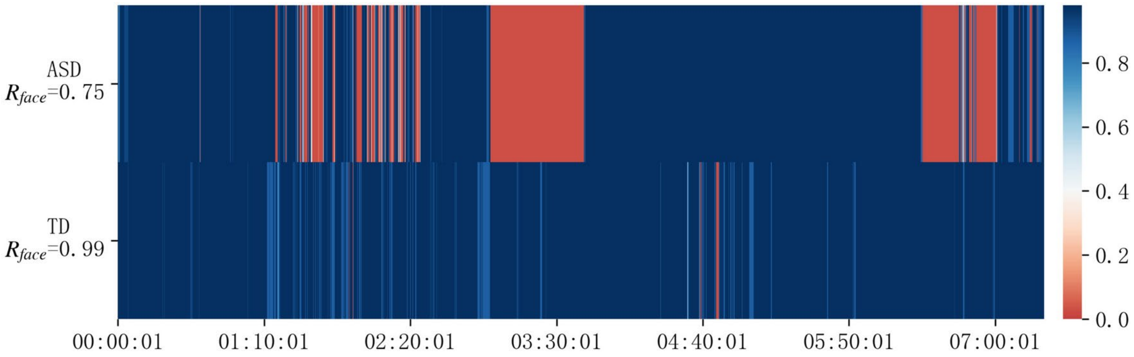
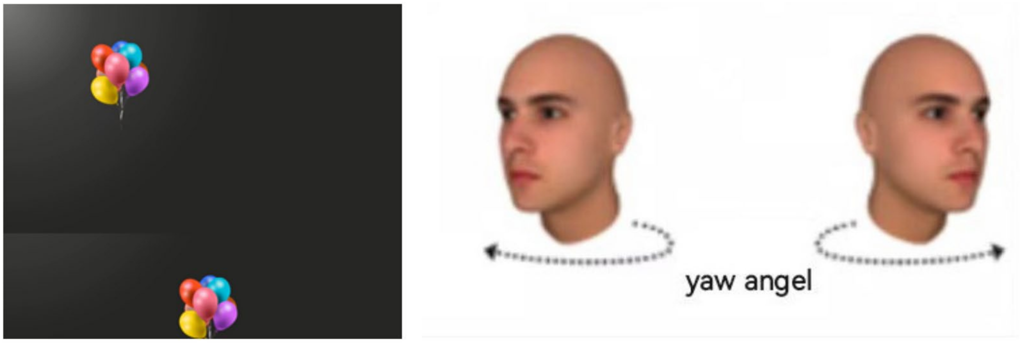


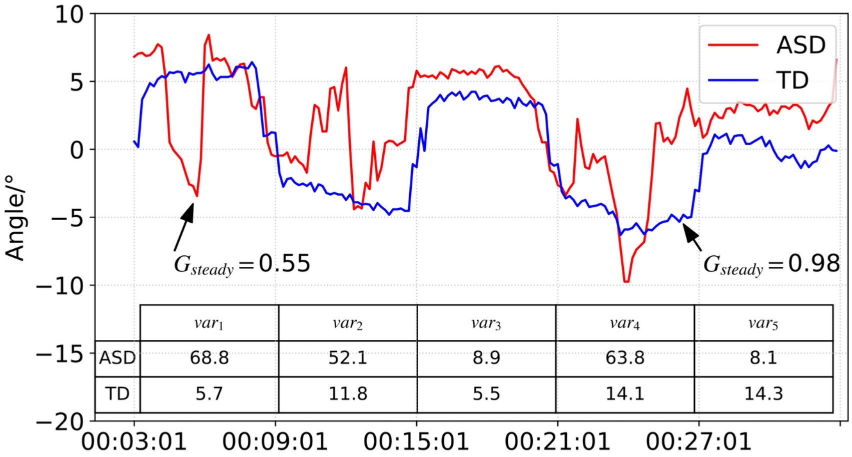
Figure 3. The comparison of facial appearing time for the children with ASD and TD. The top figure is for the child with ASD, the bottom figure is for the child with TD. The red area represents that no face is not detected and the blue area shows that a face is detected. From the figure, the involvement of the ASD group was much worse than that of the TD group. It shows that children with ASD are less attentive than children with TD.



(a)

(b)

Figure 4. The illustration of the segment F_1 and the yaw angle of the subject. (**a**) The illustration of the segment F_1 , where a ball rise. (**b**) When the ball rises, the subject's eye gaze should track the ball's rising, and the subject's yaw angle n the horizontal direction should be not change accordingly.



	var_1	var_2	var_3	var_4	var_5
ASD	68.8	52.1	8.9	63.8	8.1
TD	5.7	11.8	5.5	14.1	14.3

Figure 5. The comparison of yaw angle of 5 upward movements of the balloon on the segment F_1. The horizontal axis represents time, and the vertical axis represents the yaw angle. The red line is the yaw angle curve in ASD, and the blue line is the yaw angle curve in TD. The lines are divided into 5 groups for 5 upward movements of the balloon. For each segment, the red line is more volatile compared with the blue line, and the G_{steady} of the child with ASD is smaller than that of the child with TD. It shows that when the ball goes up, the gazes of the children with ASD sway from side to side, and the gaze stabilization ability of the children with ASD is weaker than that of the TD children.

To quantify the gaze stabilization ability of children, we process the yaw angle as follows. Note that the standard deviation reflects the dispersion of data. The smaller the standard deviation of each group of yaw angle, the smaller the numerical dispersion, and the stronger the gaze stability. For the five groups of yaw angle data, we define a standard deviation sequence $F_{var} = \{ var_1, var_2, var_3, var_4, var_5 \}$, and then normalize the sequence F_{var} by the sigmoid function flipped symmetrically along the yaxis. Finally, we compute the mean value and denote it as G_{steady}:

$$G_{steady} = \frac{\sum_{i=1}^{card(F_{\text{var}})} sigmoid\left(\text{var}_i\right)}{card\left(F_{\text{var}}\right)}, \left(\text{var}_i \in F_{\text{var}}\right) \tag{2}$$

The value of G_{steady} reflects children's gaze stabilization ability. The lower the value, the weaker the child's gaze stabilization ability. Otherwise, the stronger the child's gaze stabilization ability is.

In the experiment, an ASD child and a TD child are randomly selected to analyze the yaw angle data and the value of G_{steady}, as shown in Fig. 5. The G_{steady} of the child with ASD is 0.55 and these five groups of standard deviation data are larger. However, the G_{steady} of the TD child is 0.98 and these 5 groups of standard deviation data are all smaller. It shows that when the ball goes up, the gazes of the children with ASD swing from side to side, and the gaze stabilization ability of the children with ASD is weaker than that of the TD children.

(2).Gaze-tracking ability

The segment F_3 is a video of children observing balloons bouncing left and right. When a subject observed the ball bouncing back and forth, the subject's yaw angle continued to decrease when the ball moved to the right, and the subject's yaw angle continued to increase when the ball moved to the left. Because the yaw angle reflects the subject's gaze angle in the horizontal direction, it can analyze their gaze tracking ability. The illustration of the segment F_3 and the yaw angle of a subject is shown in the following Fig. 6.

The segment is divided into four groups according to the direction of movement of the ball. The yaw angles of the children are shown for each group in Fig. 7. In this figure, for each group, the yaw angle curve in ASD (red line) is more volatile than the yaw angle curve in TD (blue line), showing that children with ASD have a poorer gaze tracking ability than children with TD.

To quantify the gaze-tracking ability of children, we have processed the yaw angle as follows. The ball in the first group and the third group moves to the right, and 48 frames in the first group and the third group are classified as sequential F_{right}. The video frames are all facial images of the subjects watching the ball move to the right in sequence F_{right}. When the eye moves to the right, the yaw angle continues to decrease and the derivative of the yaw angle, $gaze\psi(f_i)'$, should be less than zero. Equation (3) counts the number of frames in which $gaze\psi(f_i)'$ is less than or equal to zero in the sequence F_{right}, and is denoted as G_{f1}.

$$G_{f1} = card\{(gaze\psi(f_i)' \leq 0, f_i \in F_{right})\}/card(F_{right}) \tag{3}$$

In contrast, the ball in the second and fourth groups moves to the left, and 48 frames in the second group and the fourth group are classified as the sequence F_{left}. The video frames are all facial images of the subjects watching the ball move to the left in sequence F_{left}. When the eye moves to the left, the yaw angle continues to increase and the derivative of the yaw angle, $gaze\psi(f_i)'$, should be greater than zero. Equation (4) counts the number of frames in which $gaze\psi(f_i)'$ is greater than or equal to zero in the sequence F_{left}, and is denoted as G_{f2}.

$$G_{f2} = card\{(gaze\psi(f_i)' \geq 0, f_i \in F_{left})\}/card(F_{left}) \tag{4}$$

Finally, according to (5), G_{follow}, the average value of G_{f1} and G_{f2}, is got to represent the gaze tracking ability. The smaller the value of G_{follow}, the weaker the gaze-tracking ability of the child. Otherwise, the stronger the gaze-tracking ability of the child.

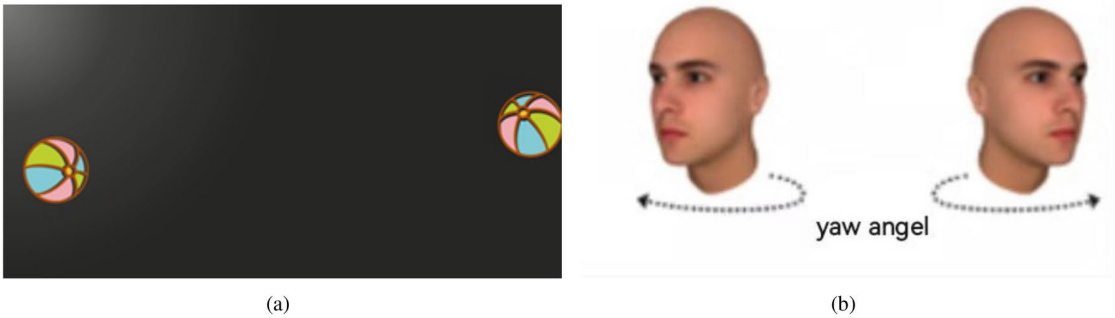


Figure 6. The illustration of the segment F_3 and the yaw angle of a subject. (**a**) The illustration of the segment F_3 , where a ball bounces back and forth. (**b**) When the ball bounces back and forth horizontally, the subject's eye gaze will track the ball's bounce in the horizontal direction, and the subject's yaw angle will also change accordingly.

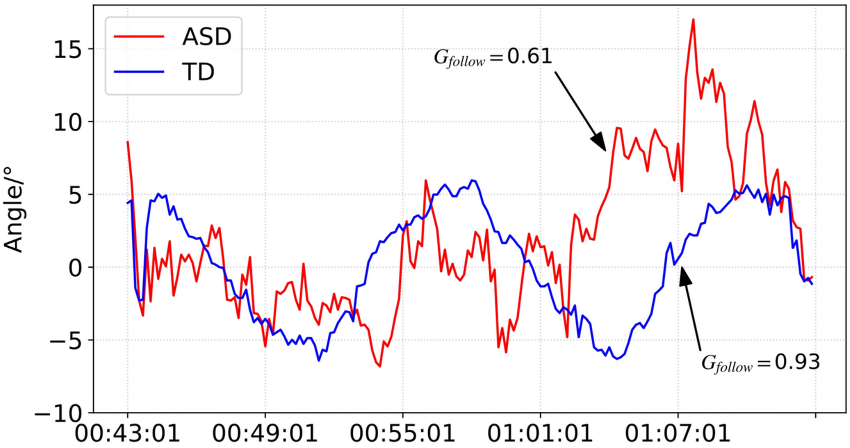


Figure 7. The comparison of variation of the yaw angle when the children watching the segment F_3. The horizontal axis represents time, and the vertical axis represents the yaw angle of the head pose. The red line shows the yaw angle curve change in ASD, and the blue line shows the yaw angle curve change in TD. From the figure, the red line is more volatile compared with the blue line, and the G_{follow} of the child with ASD is smaller than that of the child of TD, which shows that the children with ASD have poorer gaze tracking ability than the children with TD.

$$G_{follow} = (G_{f1} + G_{f2})/2 \tag{5}$$

In the experiment, a child with ASD and a child with TD are randomly selected to analyze their changes in the yaw angle in the segment F_3, as shown in Fig. 7. The value of G_{follow} in an ASD child is 0.61, while G_{follow} of a TD child is 0.93, showing that children with ASD have a poorer gaze-tracking ability than children with TD.

- name calls response time

  The segment F_2 and the segment F_4 are name-call tests. When testing, the experimenter stands on the left and right sides of the subject and calls the subject's name at the time of the call. The response time is the time it takes for the subject to turn their head in response to hearing the name. Test whether the child responds to the call and how long the response time is.By the definition of $pose\psi(f_i)$, it can be used to study the response time of the subject. Denote the time point for the calling of the name as T_0. Starting from the time point T_0, we extract the yaw angle of the head pose of each frame in 10 seconds and form a sequence $\text{F}_{pose\psi}$. The time point corresponding to the first peak value of the sequence $\text{F}_{pose\psi}$ is the reaction time point of the subject, denoting as T. The response time *t* is the difference between the response time point T and the call time point T_0, that is, $t = T - T_0$, in milliseconds. Then we normalize the response time *t* to the interval [0,1] with the sigmoid function with parameter *w* as 0.0015 and threshold value as 3500. After normalization, the response time for the first call is denoted as T_{react1} and that for the second call is denoted as T_{react2}.

The difference in this experiment between children with ASD and children with TD is shown in Fig. 8. The response time for children with ASD is 3,800ms, and the normalized value of T_{react} is 0.34, the response time of children with TD is 1,300ms, and the normalized value of T_{react} is 0.96. The longer response time and the smaller T_{react} suggest that children with ASD are less responsive to name calls.

- Happy emotion expression ability

  Smiling is the main way to express happiness on the face. According to the statistics of two facial data sets of CK+[41] and RAF-DB[42], the most important facial action units in smiling expressions are the facial action units AU6 and AU12. In the facial action unit coding system, AU6 denotes eye contraction and cheek lifting, while AU12 denotes mouth corner lifting.In this study, the sixth segment of the experimental material F_6 consists of two videos of cheerful scenes, which are easy to stimulate children to express happy emotions, lasting 5 minutes. The emotion of the children is detected and analyzed according to the intensity of the facial action unit AU6 and AU12, as seen in Fig. 9, which displays the intensity values of AU6 and AU12. The blue line, valued in [0,1], indicates the intensity of happiness. The figure shows that children with ASD have less happy emotion expressions when they see the cheerful scenes.

In order to further quantify children's emotions, we have processed the intensity values of AU6 and AU12 as follows. The piecewise function (6) is used to discretize the two variables, and the value obtained by the addition of the two variables is regarded as the feature value of happy emotion. Equations (7) and (8) are used to calculate the proportion of happy emotion when the child watches segment F_6, and the feature value S_{happy} of facial emotion detection is obtained by normalizing with the sigmoid function. The smaller the value of S_{happy} is, the

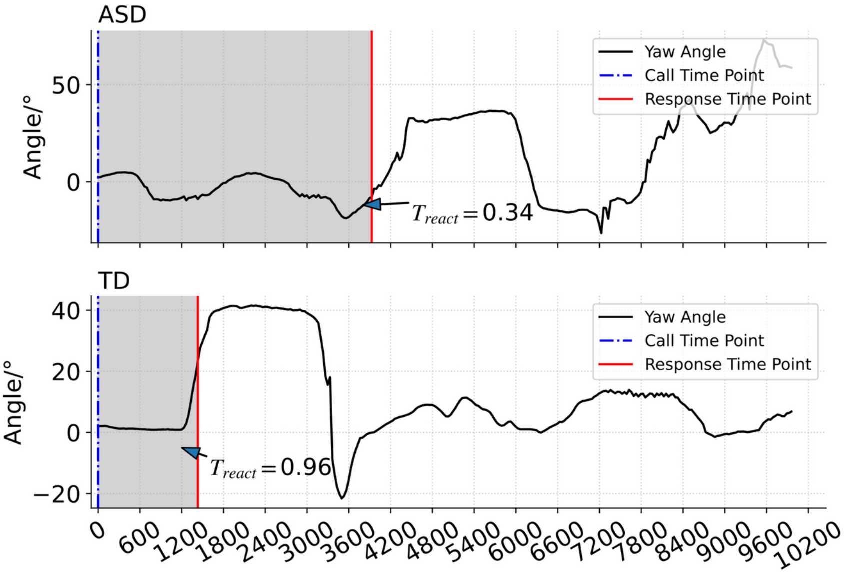


Figure 8. The comparison of name calls response time for the children with ASD and TD. The horizontal axis represents time, and the vertical axis represents the yaw angle of the head pose in the horizontal direction. The black line shows the yaw angle curve, the blue vertical line shows the call time point, and the red vertical line shows the response time point of children. From the figure, the response time of the child with ASD is longer than that of the child with TD. It shows that children with ASD are less responsive to name calls.

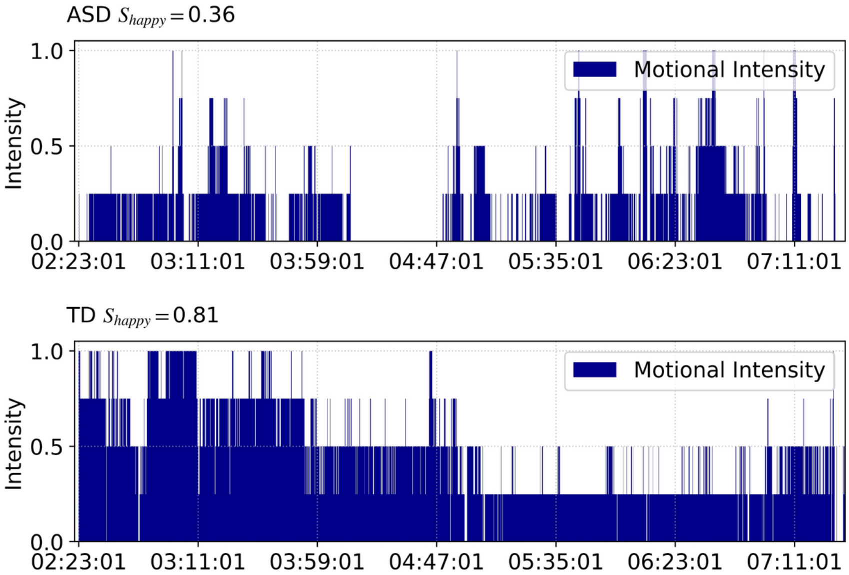


Figure 9. The intensity values of AU6 and AU12. The top figure is for the child with ASD, the bottom figure is for the child with TD. The horizontal axis represents time, and the vertical axis represents the happy intensity value. The blue are represents happy intensities. The figure shows that children with ASD have less happy emotion expressions.

weaker the children's ability to express happy emotion is, otherwise, the stronger their ability to express happy emotion is.

$$dis(x)=\begin{cases}0 & x=0\\ 0.25 & 0<x<0.5\\ 0.5 & 0.5\leq x\leq 1\end{cases}\tag{6}$$

$$r=\frac{\sum_{i=0}^{card(F_6)} dis\left(AU_6\left(f_i\right)\right)+dis\left(AU_{12}\left(f_i\right)\right)}{card\left(F_6\right)}\tag{7}$$

$$S_{happy} = sigmoid\left(r\right) \tag{8}$$

From Fig. 9, the proportion of happy emotions in children with ASD is low and the S_{happy} value is small, showing that children with ASD have a weaker ability to express happy emotions. The proportion of happy emotions in children with TD is high, and the S_{happy} value is larger, showing that children with TD have a stronger ability to express happy emotions.

Data availability

Other researchers can access this dataset, provided they have signed a confidentiality agreement before use. The agreement requires those users must comply with corresponding ethics when using this dataset, and can only use the dataset for scientific research, and is prohibited from using it for any commercial activities. Furthermore, the confidentiality agreement imposes a duty on users to protect the personal privacy of children and prohibits the disclosure of their data. Based on computer vision technology, we have also developed a dataset of digital features for children, which can be used to describe typical symptoms of autism and will be publicly used in scientific research. Researchers interested in this data set can contact the first author Ruisheng Ran.

Received: 3 March 2024; Accepted: 19 November 2024
Published online: 23 November 2024

References

1. Association, A. P. *Diagnostic and statistical manual of mental disorders DSM-5*. vol. 5 (American psychiatric association Washington, DC, 2013).
2. Lord C., B. T. S. Autism spectrum disorder. *Nat. Rev. Dis. Primers.* **6**(1), 1–23. https://doi.org/10.1038/s41572-019-0138-4 (2020).
3. Naviaux, R. K. et al. *Infantile Autism: The Syndrome and Its Implications for a Neural Theory of Behavior by Bernard Rimland*. Ph.D. thesis, Jessica Kingsley Publishers (2014).
4. Clancy, H., Dugdalei, A. & Rendle-Shortt, J. The diagnosis of infantile autism. *Dev. Med. Child Neurol.* **11**, 432–442. https://doi.org/10.1111/j.1469-8749.1969.tb01461.x (1969).
5. Krug, D. A., Arick, J. & Almond, P. Behavior checklist for identifying severely handicapped individ-uals with high levels of autistic behavior. *Child Psychol. Psychiatry Allied Discip.* **21**, 221–229. https://doi.org/10.1111/j.1469-7610.1980.tb01797.x (1980).
6. Lord, C. et al. Ados.autism diagnostic observation schedule. *Manual. Los Angeles: WPS*. https://doi.org/10.1111/j.1469-7610.1980.tb01797.x (1999).
7. Baron-Cohen, S., Allen, J. & Gillberg, C. Can autism be detected at 18 months? the needle, the haystack, and the chat. *Br. J. Psychiatry* **161**, 839–843. https://doi.org/10.1192/bjp.161.6.839 (1992).
8. Hashemi, J. et al. Computer vision analysis for quantification of autism risk behaviors. *IEEE Trans. Affect. Comput.* **12**, 215–226. https://doi.org/10.1109/TAFFC.2018.2868196 (2018).
9. Zwaigenbaum, L. et al. Behavioral manifestations of autism in the first year of life. *Int. J. Dev. Neurosci.* **23**, 143–152. https://doi.org/10.1016/j.ijdevneu.2004.05.001 (2005).
10. Kohli, M., Kar, A. K. & Sinha, S. The role of intelligent technologies in early detection of autism spectrum disorder (asd): A scoping review. *IEEE Access* **10**, 104887–104913. https://doi.org/10.1109/ACCESS.2022.3208587 (2022).
11. Manfredonia, J. et al. Automatic recognition of posed facial expression of emotion in individuals with autism spectrum disorder. *J. Autism Dev. Disord.* **49**, 279–293. https://doi.org/10.1007/s10803-018-3757-9 (2019).
12. Egger, H. L. et al. Automatic emotion and attention analysis of young children at home: a researchkit autism feasibility study. *NPJ Digit. Med.* **1**, 1–10. https://doi.org/10.1038/s41746-018-0024-6 (2018).
13. Paolucci, C., Giorgini, F., Scheda, R., Alessi, F. V. & Diciotti, S. Early prediction of autism spectrum disorders through interaction analysis in home videos and explainable artificial intelligence. *Comput. Hum. Behav.* **148**, 107877. https://doi.org/10.1016/j.chb.2023.107877 (2023).
14. Pierce, K. et al. Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder sub-type associated with increased symptom severity. *Biol. Psychiat.* **79**, 657–666. https://doi.org/10.1016/j.biopsych.2015.03.032 (2016).
15. Bovery, M., Dawson, G., Hashemi, J. & Sapiro, G. A scalable off-the-shelf framework for measuring patterns of attention in young children and its application in autism spectrum disorder. *IEEE Trans. Affect. Comput.* **12**, 722–731. https://doi.org/10.1109/TAFFC.2018.2890610 (2019).
16. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis. *Sci. Rep.* **10**, 1–13. https://doi.org/10.1038/s41598-020-65384-4 (2020).
17. Alvari, G., Coviello, L. & Furlanello, C. Eye-c: Eye-contact robust detection and analysis during un-constrained child-therapist interactions in the clinical setting of autism spectrum disorders. *Brain Sci.* **11**, 1555. https://doi.org/10.3390/brainsci11121555 (2021).
18. Wei, Q. et al. Early identification of autism spectrum disorder based on machine learning with eye-tracking data. *J. Affect. Disord.* **358**, 326–334. https://doi.org/10.1016/j.jad.2024.04.049 (2024).
19. Martin, K. B. et al. Objective measurement of head movement differences in children with and without autism spectrum disorder. *Mol. Autism.* **9**, 1–10. https://doi.org/10.1186/s13229-018-0198-4 (2018).
20. Dawson, G. et al. Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. *Sci. Rep.* **8**, 1–7. https://doi.org/10.1038/s41598-018-35215-8 (2018).
21. Zhao, Z. et al. Identifying autism with head movement features by implementing machine learning algorithms. *J. Autism Dev. Disord.* 1–12. https://doi.org/10.1007/s10803-021-05179-2 (2022).
22. Nakai, Y., Takiguchi, T., Matsui, G., Yamaoka, N. & Takada, S. Detecting abnormal word utterances in children with autism spectrum disorders: machine learning-based voice analysis versus speech therapists. *Percept. Mot. Skills* **124**, 961–973. https://doi.org/10.1177/0031512517716855 (2017).
23. Ramesh, V. & Assaf, R. Detecting autism spectrum disorders with machine learning models using speech transcripts. *arXiv preprint* . arXiv:2110.03281. https://doi.org/10.48550/arXiv.2110.03281 (2021).
24. Hashemi, J. et al. Computer vision tools for low-cost and noninvasive measurement of autism related behaviors in infants. *Autism Res. Treatm.* **2014**. https://doi.org/10.1155/2014/935686 (2014).
25. Jiang, M. et al. Classifying individuals with asd through facial emotion recognition and eye-tracking. *In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 6063–6068, https://doi.org/10.1109/EMBC.2019.8857005 (2019).

26. Jaiswal, S., Valstar, M. F., Gillott, A. & Daley, D. Automatic detection of adhd and asd from expressive behaviour in rgbd data. *In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition*. 762–769, https://doi.org/10.1109/FG.2017.95 (2017).
27. Li, J., Chen, Z., Li, G., Ouyang, G. & Li, X. Automatic classification of asd children using appearance-based features from videos. *Neurocomputing* **470**, 40–50. https://doi.org/10.1016/j.neucom.2021.10.074 (2022).
28. Lubetsky, M. J. & Handen, B. Medication treatment in autism spectrum disorder. *Speaker's J.* **8**, 97–107 (2008).
29. Ali, A., Negin, F. F., Bremond, F. F. & Thümmler, S. Video-based behavior understanding of children for objective diagnosis of autism. In *VISAPP 2022-17th International Conference on Computer Vision Theory and Applications*, https://doi.org/10.5220/0010839200003124 (2022).
30. Varghese, E. B., Qaraqe, M., Al Thani, D. & Ekenel, H. K. Attention assessment in children with autism using head pose and motion parameters from real videos. In *GLOBECOM 2023-2023 IEEE Global Communications Conference*, 6462–6468, https://doi.org/10.1109/GLOBECOM54140.2023.10436851 (2023).
31. Ball, R. S. The gesell developmental schedules: Arnold gesell (1880–1961). *J. Abnorm. Child Psychol.* **5**, 233–239. https://doi.org/10.1007/BF00913694 (1977).
32. Robins, D. L., Fein, D., Barton, M. L. & Green, J. A. The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J. Autism Dev. Disord.* **31**, 131–144. https://doi.org/10.1023/A:1010738829569 (2001).
33. Wong, V. et al. A modified screening tool for autism (checklist for autism in toddlers [chat-23]) for Chinese children. *Pediatrics* **114**, e166–e176. https://doi.org/10.1542/peds.114.2.e166 (2004).
34. Lord, C. et al. The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30**, 205–223. https://doi.org/10.1023/A:1005592401947 (2000).
35. Baltrusaitis, T., Zadeh, A., Lim, Y. C. & Morency, L. P. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, 59–66, https://doi.org/10.1109/FG.2018.00019 (IEEE, 2018).
36. Zadeh, A., Chong Lim, Y., Baltrusaitis, T. & Morency, L. P. Convolutional experts constrained local model for 3d facial landmark detection. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2519–2528. https://doi.org/10.48550/arXiv.1611.08657 (2017).
37. Hesch, J. A. & Roumeliotis, S. I. A direct least-squares (dls) method for pnp. In *2011 International Conference on Computer Vision*, 383–390. https://doi.org/10.1109/ICCV.2011.6126266 (IEEE, 2011).
38. Wood, E. et al. Rendering of eyes for eye-shape registration and gaze estimation. In *Proceedings of the IEEE international conference on computer vision*, 3756–3764, https://doi.org/10.48550/arXiv.1505.05916 (2015).
39. Ekman, P. & Friesen, W. V. Facial action coding system. *Environ. Psychol. Nonverbal Behav.*[SPACE]https://doi.org/10.1037/t27734-000 (1978).
40. Baltrušaitis, T., Mahmoud, M. & Robinson, P. Cross-dataset learning and person-specific normalisation for automatic action unit detection. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 6, 1–6, https://doi.org/10.1109/FG.2015.7284869 (IEEE, 2015).
41. Lucey, P. et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE computer society conference on computer vision and pattern recognition-workshops*, 94–101, https://doi.org/10.1109/CVPRW.2010.5543262 (IEEE, 2010).
42. Li, S., Deng, W. & Du, J. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2852–2861, https://doi.org/10.1109/CVPR.2017.277 (2017).

Acknowledgements

This work was supported by Humanities and Social Sciences Project of the Ministry of Education of China under grant 20YJAZH084. The humanities and social sciences research project of Chongqing Municipal Commission of education in 2024, "Research on the innovation and application path of diagnosis mode of autism in children from the perspective of artificial intelligence (24SKGH065)".

Author contributions

R. R. proposes the idea, writing, and funding. W. L. presents the method, experiments, writing of the original draft. S. D. responsible for writing and submission, X. F. provides the experimental environment and recruits children, and project administration. C. L. designs the data collection scence. W. L. and Q. H. design and produce the experimental material. K. S., T. W., S. D., and Q. H. collect the videos of children.

Declarations

Competing interests

The authors declare no competing interests.

Statement

This research was prospectively reviewed and approved by the Medical Ethics Committee of Chongqing Maternal and Child Health Hospital. This research does not involve clinical trials. In the data collection stage, we have children watch experimental materials (some prepared short videos) in front of a computer screen, and record the videos of children watching the materials. Then, some computer vision methods are used for the analysis of the videos of children, so no clinical trials are conducted.

Additional information

Correspondence and requests for materials should be addressed to X.F.