Appendix

Participants. The sample size for this study was calculated using G * Power3.1 software[6], A mixed-design repeated measures ANOVA was used, with the significance level $\alpha = 0.05$, statistical power $(1-\beta) = 0.80$, effect size medium ($f = 0.25$). A total of 34 participants are needed. After excluding 6 pairs, 29 autistic children and 28 nonautistic children were left. Demographic characteristics for children and mothers are provided in Appendix Table 1.

Inclusion criteria for autistic children included: (i) a hospital-confirmed diagnosis of autism and diagnostic criteria based on the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition[2](DSM-V); (ii) the ability to follow the main test instructions(e.g., to"look at your mother's face"); (iii) no serious behavioral problems(e.g., inability to remain seated for 30 minutes or difficulty wearing a hat); (iv) no attention deficits or other neurological disorders. Exclusion criteria for nonautistic children were (i) history of neurological or developmental disorders, premature birth, or birth difficulties; (ii) use of neurological or psychotropic medications; (iii) uncorrected visual or hearing impairment; (iv) family history of autism.The exclusion criteria for mothers were consistent with those for nonautistic children, including.

Appendix Table 1. Valid demographic data of the participants

Features	autistic children ($n = 29$) $M \pm SD$	nonautistic children ($n = 28$) $M \pm SD$	Difference test
Sex	6 females 23male	12females 16male	$\chi^2(1) = 3.24$, $p = 0.072$
Age	5.23 ± 0.93	4.90 ± 0.79	$t(55) = 1.43$, $p = 0.159$, *cohen's d* = 0.39
Age(mother)	35.72 ± 4.91	36.79 ± 4.79	$t(55) = -0.83$, $p = 0.413$, *cohen's d* = −0.22
intelligence	59.89 ± 16.98	112.39 ± 21.19	$t(55) = -10.34$, $p < 0.001$, *cohen's d* = −2.79
ABC	69.48 ± 37.37	8.04 ± 10.00	$t(32) = 8.54$, $p < 0.001$, *cohen's d* = 3.02
SRS-2	101.93 ± 30.68	44.18 ± 16.20	$t(43) = 8.93$, $p < 0.001$, *cohen's d* = 2.72

Note: ABC: Autism Behavior Checklist; SRS-2: Second Edition of the Social Response Scale; The difference test includes the independent sample t-test and the chi-square test.

Materials.

(1)Questionnaire

Peabody Picture Vocabulary Test-Revised(PPVT-R): PPVT-R was developed by American psychologist L. M. Dunn[5]. It is commonly used intelligence assessment tool by the American Association on Intellectual and Development Disability. This experiment adopted the revised Chinese version by Sang Biao and Miao Xiaochun. The test contains vocabulary types such as verbs, nouns, adjectives, etc., in the form of a task of listening to words and pointing to pictures, with a total of 175 questions, which is applicable to children aged 3.5 years to 9 years and 2 months, and is currently commonly used as an intelligence measure for autistic children[1].

Social Response Scale-Second Edition (SRS-2): It was a core assessment tool for the social interaction ability of children with autism, was developed by the Constantino research team[4]. This system conducts standardized evaluations on the social interaction, non-verbal communication and stereotyped behaviors of individuals aged 48 to 216 months through a parent/teacher proxy reporting model. This assessment system consists of 65 structured items and adopts the Likert four-point scoring method. The cumulative score is positively correlated with the degree of social dysfunction. The Chinese version was verified for cross-cultural validity by the team of Gong Jun (2019)[7], and its good reliability and validity provide a reliable tool for clinical assessment.

Autism Behavior Checklist (ABC): It was initially developed by Krug in 1980[9]. It covers five major symptoms: Sensory, Relating, Body concept, Language, Self-care, with a total of 57 clinical observation indicators. The Chinese revised version was localized and verified by Professor Yang Xiaoling's team[10]. Its standardized scoring system adopts a symptom weight scoring mechanism (with a gradient of 1-4 points), and a cumulative total score of $\geq$ 67 points has clinical diagnostic reference value. This tool is particularly suitable for symptom screening in the autistic population aged 18 months to school age. Its multi-dimensional symptom assessment framework has established good ecological validity in clinical practice in China.

Self-Assessment Manikin(SAM): It rated the overall emotional pleasure and

arousal of individuals on a 9-point scale[3]. 1 point represents very unhappy/very calm, and 9 points represents very happy/very excited.

Self-reported 16 emotion items: It used by Gross and Levenson(1995)[8] when evaluating film clips was adopted. This questionnaire contains 6 positive emotions (such as happiness, interest, etc.) and 10 negative emotions (such as anger, sadness, etc.), which are used to rate the specific emotional intensity of the subjects on a 9-point scale. Among them, 1 point indicates almost no experience of this emotion, and 9 points indicates that this emotional experience is extremely intense. Through this questionnaire, it is possible to test whether the subjects have experienced emotional responses unrelated to the experiment.

(2) Videos

Before the experiment, the experimenter selected eight 3-minute videos with different emotional contents from publicly available platforms. They are consisted of two happy, two sad, and four neutral videos (as the baseline). To validate these stimuli, 40 female college students (age: $M = 23.5 \pm 2.03$years) were recruited rate each video on its pleasure (valence) and arousal levels using a 9-point scale. The ratings for the two happy and two sad videos were compared using paired-sample *t*-tests to ensure consistency within each emotional category. A one-way repeated-measures ANOVA was then conducted on the four neutral videos to confirm they did not differ significantly from one another in their emotional impact.

The results showed that there was no significant difference between the pleasure of the two happy videos, t (39) = 0.71, p = 0.482, *cohen's d* = 0.23. There was no significant difference in arousal between the two happy videos, t (39) = 0.36, p = 0.720, *cohen's d* = 0.16. There was no significant difference in the pleasure between the two sad videos, t (39) = 0.50, p = 0.623, *cohen's d* = 0.16. There was no significant difference in arousal between the two sad videos, t (39) = 0.57, p = 0.570, *cohen's d* = 0.18. The main effect of the pleasure of the four neutral videos was significant, F (3, 117) = 0.64，p = 0.590，η_p^2 = 0.02. Post hoc tests showed that there was no significant difference in pleasure between each pair of videos(ps ≥ 0.173). The

main effect of arousal for the four neutral videos was significant, F (3, 117) = 1.26， p = 0.291， η_p^2 = 0.03. Post hoc tests showed that there was no significant difference in titers between each pair of videos(ps ≥ 0.183).

Further one-way repeated measures ANOVA was conducted on the pleasure and arousal of three emotional videos types(happy verse sad verse neutral). The results showed that main effect of the pleasure for the video emotional type was significant, F (2, 78) = 1749.70， p < 0.001， η_p^2 = 0.98. Further post hoc analysis confirmed that there were significant differences in the pleasure of each pair of the three types of emotional videos(ps < 0.05). The main effect of the arousal for the video emotion type was significant, F (2, 78) = 3028.66，p < 0.001，η_p^2 = 0.98. Post hoc analysis indicated that there were significant differences in the arousal of each pair of the three types of emotion videos(ps < 0.05)(see Appendix Table 2 for details).

Appendix Table 2. Video Valence and Arousal

Video type	Valence($M \pm SD$)	Arousal($M \pm SD$)
Happy 1	8.23 ± 0.62	8.30 ± 0.72
Happy 2	8.10 ± 0.74	8.25 ± 0.78
Neutral 1	4.98 ± 0.66	1.10 ± 0.30
Neutral 2	5.18 ± 0.68	1.25 ± 0.59
Neutral 3	5.10 ± 0.68	1.12 ± 0.33
Neutral 4	5.05 ± 0.78	1.10 ± 0.30
Sad 1	1.45 ± 0.75	8.08 ± 0.86
Sad 2	1.40 ± 0.71	8.03 ± 0.80

Machine Learning Methods

1.1Classifier

The dataset in this study contains a limited number of samples. To address this challenge, we employed decision trees as our classification model. As a tree-structured supervised learning algorithm, decision trees perform classification through a series of "if-then" rules by recursively splitting data from the root node based on feature values until reaching leaf nodes. The selection of decision trees was motivated by three key considerations: (1) Decision trees demonstrate superior performance to deep learning algorithms in small-sample scenarios; (2) Their decision-making process provides intuitive and interpretable results that facilitate

subsequent analysis; (3) They enable direct evaluation of feature importance, thereby simplifying subsequent feature selection procedures.

The decision tree was configured with the following parameters: the CART (Classification and Regression Trees) algorithm was employed, using the Gini coefficient as the splitting criterion, with the maximum tree depth set to 5, the minimum samples for node splitting set to 2, and the minimum samples required at leaf nodes set to 1.

1.2 Feature Selection

The functional connectivity data and inter-brain synchronization data contain a relatively large number of features, with 132 and 288 features respectively. Directly inputting these features into the classifier model would result in suboptimal classification performance, as the numerous redundant features would impair the model's ability to learn meaningful patterns. To address this issue, this study employs a wrapper method for feature selection on these two datasets.

The wrapper method evaluates features by utilizing the classifier model itself. It trains the classifier on the data to compute feature importance scores, thereby retaining the most informative features while eliminating less important ones. To prevent overfitting, all experiments employ 5-fold cross-validation. Furthermore, to determine the optimal number of features, we conduct exhaustive experiments testing feature subset sizes ranging from 1 to 20. Given the inherent stochasticity of wrapper methods, each experiment is repeated 30 times, with the final results averaged across all repetitions. The specific algorithm implementation is detailed below:

Input: Dataset $S = \{f_1, f_2, \ldots, f_n\}$, where f_i represents the i-th feature. $|S|$ denotes the total number of features. K represents the maximum number of features to select. R is the number of experimental repetitions

Output: Optimal feature subset S_{best}

Procedure:

1. For each feature number $k \in [1, 2, ..., K]$ do:
    1.1 For each repetition $r \in [1, 2, ..., R]$ do:
        1.1.1 Initialize current feature subset: $S_r \leftarrow S$
        1.1.2 While $|S_r| > k$ do:

1.1.2.1 Train classifier model using S_r

1.1.2.2 Compute feature importance for each feature f_i

1.1.2.3 Remove the least important feature f_{min} from S_r

1.1.3 Evaluate classifier on S_r using 5-fold cross-validation, obtain performance P_r

1.2 Compute the average performance P_k and best performance P_{max} across R repetitions: $P_k \leftarrow \text{mean}(P_r)$, $P_{max} \leftarrow \text{argmax}(P_r)$,

1.3 Obtain the best feature subset S_k corresponding the best performance P_{max}.

2. Determine optimal feature number: $P_{best} \leftarrow \text{argmax}(P_k)$

3. Obtain the best feature subset S_{best} corresponding to the best performance P_{best}

1.3 Multimodal Ensemble Learning

For each of the four modalities in our dataset, we constructed individual decision tree classifiers. The facial expression and heart rate data were classified directly using their raw features, while the functional connectivity and inter-brain synchronization data were first preprocessed through feature selection before being fed into their respective classifiers.

To effectively integrate information across all modalities, we implemented an ensemble approach that averages the output probabilities from each decision tree. We evaluated all 15 possible modality combinations through exhaustive enumeration. In this framework, each modality's classifier is assigned a binary weight ($c_i \in \{0,1\}$), where $c_i = 1$ indicates inclusion in the ensemble and $c_i = 0$ indicates exclusion. The ensemble's final output is computed as:

$$y_{ensemble} = \sum_{i=1}^{4} c_i y_i$$

Machine Learning of Feature Selection Results

Through our wrapper-based feature selection approach (Section 1.2), we identified optimal feature subsets consisting of 11 key features for functional connectivity data (representing 8.3% of the original features) and 7 critical features for inter-brain synchronization data (representing 2.4% of the original features). The selected features are presented in Appendix Tables 3 and 4.

Appendix Table 3. Selected Features for Functional Connectivity Data

Feature Number	Feature Decription
8	Happy ROI1-9
9	Happy ROI1-10
44	Happy ROI5-11
73	Sad ROI1-8
98	Sad ROI4-6
99	Sad ROI4-7
101	Sad ROI4-9
102	Sad ROI4-10
105	Sad ROI5-6
131	Sad ROI10-12
132	Sad ROI11-12

Appendix Table 4. Selected Features for Inter-Brain Synchronization Data

Feature Number	Feature Description
108	Happy ROI9-12
153	Sad ROI1-9
160	Sad ROI2-4
247	Sad ROI9-7
252	Sad ROI9-12
274	Sad ROI11-10
275	Sad ROI11-11

From the tables, we can see that in both datasets, the number of features corresponding to sad emotions is greater than that for happy emotions. For the functional connectivity data, there are 8 features associated with sad emotions, compared to only 3 features associated with happy emotions. For the inter-brain synchronization data, the difference is even more pronounced, with 6 features corresponding to sad emotions versus merely 1 feature corresponding to happy emotions. This indicates that, from the classifier's perspective, sad emotions exhibit greater discriminability than happy emotions between the two groups of children.

Regarding the brain regions associated with these features, the features are not uniformly distributed across all brain regions but are instead relatively concentrated in a few specific regions. For the functional connectivity data, the primary corresponding brain regions are: region 4 (4 features), region 10 (3 features), and

region 1 (3 features). For the inter-brain synchronization data, the main corresponding brain regions are: region 9 (4 features), region 11 (2 features, appearing in 3 connections), and region 12 (2 features)

Appendix Table 5. The Subjective Emotional Valence and Arousal of the Emotional Sender

Type	Happy($M \pm SD$)	Sad($M \pm SD$)	Neutral($M \pm SD$)
Valence	7.97 ± 0.94	1.75 ± 0.68	5.10 ± 0.19
Arousal	7.08 ± 1.42	5.33 ± 2.33	1.71 ± 0.79

Appendix Table 6. The Heart Rate

Type	Happy($M \pm SD$)	Sad($M \pm SD$)	Neutral($M \pm SD$)
Emotional sender	72.32 ± 10.92	70.51 ± 8.78	67.42 ± 8.91
Emotional receiver	81.88 ± 15.24	86.01 ± 14.53	73.68 ± 11.79

Appendix Table 7. Descriptive Statistics of Facial Expressions and Heart Rate

Type		Happy($M \pm SD$)	Sad($M \pm SD$)
The facial expression intensity	autistic-sender	0.79 ± 0.55	0.12 ± 0.06
	nonautistic-sender	0.93 ± 0.49	0.15 ± 0.13
	autistic-receiver	0.17 ± 0.10	0.39 ± 0.13
	nonautistic-receiver	0.19 ± 0.24	0.22 ± 0.10
The facial expression synchrony	autistic dyadic group	0.07 ± 0.11	0.09 ± 0.07
	nonautistic dyadic group	0.16 ± 0.15	0.13 ± 0.12
The imitation imprecision score	autistic dyadic group	0.75 ± 0.47	0.44 ± 0.14
	nonautistic dyadic group	0.86 ± 0.44	0.31 ± 0.15
The heart rate	autistic-sender	73.32 ± 12.18	70.14 ± 9.62
	nonautistic-sender	71.28 ± 9.56	70.89 ± 7.97
	autistic-receiver	73.26 ± 12.56	80.22 ± 14.99
	nonautistic-receiver	90.81 ± 12.49	92.00 ± 11.48
The heart rate synchrony	autistic dyadic group	0.03 ± 0.36	0.09 ± 0.30
	nonautistic dyadic group	0.13 ± 0.25	-0.01 ± 0.37

Appendix Table 8. Descriptive Statistics of Near-infrared Data

Type		Happy($M \pm SD$)	Sad($M \pm SD$)
ROI3(rIFG)-ROI11(rPMC)	autistic-sender	-0.015 ± 0.05	0.033 ± 0.04
functional connectivity value	nonautistic-sender	0.009 ± 0.06	0.030 ± 0.07
ROI6(lSTS)-ROI8(lIPL)	autistic-sender	0.013 ± 0.04	0.018 ± 0.03
functional connectivity value	nonautistic-sender	0.029 ± 0.03	-0.013 ± 0.05
ROI5(rSTS)-ROI11(rPMC)	autistic-sender	-0.024 ± 0.07	0.033 ± 0.05
functional connectivity value	nonautistic-sender	0.013 ± 0.07	0.029 ± 0.05
ROI9(rPSC)-ROI11(rPMC)	autistic-receiver	-0.013 ± 0.06	0.042 ± 0.09

functional connectivity value	nonautistic-receiver	0.026 ± 0.07	-0.022 ± 0.08
$Sender_{ROI7(rIPL)}$-$Receiver_{ROI9(rPSC)}$ IBS	autistic dyadic group	0.011 ± 0.02	0.023 ± 0.03
	nonautistic dyadic group	0.022 ± 0.03	0.005 ± 0.02
$Sender_{ROI8(lIPL)}$-$Receiver_{ROI9(rPSC)}$ IBS	autistic dyadic group	0.007 ± 0.03	0.025 ± 0.03
	nonautistic dyadic group	0.021 ± 0.02	0.003 ± 0.03
Receiver→Sender ROI8-ROI19 causality value	autistic dyadic group	0.087 ± 0.03	0.068 ± 0.02
	nonautistic dyadic group	0.068 ± 0.02	0.076 ± 0.03
Sender→Receiver ROI8-ROI19 causality value	autistic dyadic group	0.066 ± 0.02	0.074 ± 0.02
	nonautistic dyadic group	0.075 ± 0.02	0.078 ± 0.04

Appendix Table 9. MNI Coordinates

Channel	MNI(mm)			BA	Anatomical Labels	Percentage
	x	y	z			
1	71.96	-34.14	6.71	22	Superior Temporal Gyrus	0.74
2	72.02	-19.68	-11.18	21	Middle Temporal gyrus	0.95
3	69.69	-20.15	28.63	2	Primary Somatosensory Cortex	0.69
4	68.54	-7.75	9.90	22	Superior Temporal Gyrus	0.60
5	64.96	5.73	27.86	6	Pre-Motor and Supplementary Motor Cortex	0.59
6	63.22	-55.49	37.95	40	Supramarginal gyrus part of Wernicke's area	0.47
7	67.08	-33.81	43.66	40	Supramarginal gyrus part of Wernicke's area	0.73
8	63.01	4.35	-10.94	21	Middle Temporal gyrus	0.66
9	60.86	24.38	10.94	45	pars triangularis Broca's area	0.62
10	56.02	40.57	-4.96	45	pars triangularis Broca's area	0.44
11	56.32	-56.60	50.44	40	Supramarginal gyrus part of Wernicke's area	0.56
12	43.60	-56.03	60.26	40	Supramarginal gyrus part of Wernicke's area	0.63
13	59.81	-34.37	54.71	40	Supramarginal gyrus part of Wernicke's area	0.60
14	47.55	-32.55	66.48	3	Primary Somatosensory Cortex	0.45
15	53.89	33.08	27.90	45	pars triangularis Broca's area	0.96
16	49.43	48.04	12.62	46	Dorsolateral prefrontal cortex	0.62
17	38.38	51.71	28.78	46	Dorsolateral prefrontal cortex	0.90
18	35.59	-35.25	72.97	4	Primary Motor Cortex	0.42
19	26.14	48.02	42.72	9	Dorsolateral prefrontal cortex	0.96
20	13.83	39.92	55.56	8	Includes Frontal eye fields	0.62
21	42.80	60.53	-2.72	10	Frontopolar area	0.47

22	30.81	65.17	15.25	10	Frontopolar area	0.89
23	18.36	72.35	-0.58	10	Frontopolar area	0.54
24	15.23	61.39	32.17	10	Frontopolar area	0.55
25	1.85	51.57	45.69	9	Dorsolateral prefrontal cortex	0.98
26	3.19	68.27	17.95	10	Frontopolar area	1
27	-12.11	62.02	33.68	10	Frontopolar area	0.53
28	-11.80	40.09	56.49	8	Includes Frontal eye fields	0.65
29	-23.92	49.46	43.54	9	Dorsolateral prefrontal cortex	0.95
30	-13.43	73.28	0.28	10	Frontopolar area	0.66
31	-26.81	66.60	17.14	10	Frontopolar area	0.86
32	-40.81	62.24	-0.36	10	Frontopolar area	0.65
33	-34.52	-37.71	72.38	1	Primary Somatosensory Cortex	0.29
34	-35.30	51.50	30.62	46	Dorsolateral prefrontal cortex	0.93
35	-48.38	48.00	16.17	45	pars triangularis Broca's area	0.54
36	-51.68	34.22	31.50	45	pars triangularis Broca's area	0.88
37	-43.12	-58.54	59.62	40	Supramarginal gyrus part of Wernicke's area	0.55
38	-55.01	-60.34	50.09	39	Angular gyrus_ part of Wernicke's area	0.61
39	-45.73	-36.34	66.54	2	Primary Somatosensory Cortex	0.33
				3	Primary Somatosensory Cortex	0.33
40	-58.59	-37.39	55.97	40	Supramarginal gyrus part of Wernicke's area	0.75
41	-56.32	40.45	-2.43	45	pars triangularis Broca's area	0.64
42	-60.68	25.39	13.29	45	pars triangularis Broca's area	0.67
43	-63.00	6.37	-7.86	21	Middle Temporal gyrus	046
44	-62.24	-58.68	35.98	39	Angular gyrus_ part of Wernicke's area	0.63
45	-66.65	-35.21	43.94	40	Supramarginal gyrus part of Wernicke's area	0.75
46	-64.44	6.91	28.47	6	Pre-Motor and Supplementary Motor Cortex	0.60
47	-68.47	-5.54	11.89	22	Superior Temporal Gyrus	0.41
48	-69.47	-19.51	27.05	2	Primary Somatosensory Cortex	0.56
49	-71.60	-19.58	-9.98	21	Middle Temporal gyrus	0.94
50	-71.65	-35.89	6.57	22	Superior Temporal Gyrus	0.78

References

1. Ansuini, C., Podda, J., Battaglia, F. M., Veneselli, E., & Becchio, C. One hand, two hands, two people: Prospective sensorimotor control in children with autism. *Developmental Cognitive Neuroscience,* **29**, 86-96 (2017).
2. Association Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. American Psychiatric Publishing. (2013).
3. Bradley, M. M., & Lang, P. J. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry,* **25**(1), 49-59 (1994).
4. Constantino, J., & Gruber, C. Social responsiveness scale, second edition (SRS-2). (2012).
5. Dunn, L. M., & Dunn, L. M. *PPVT-III: Peabody Picture Vocabulary Test, ed*. American Guidance Service, Circle Pines. (1997).
6. Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods,* **39**(2), 175-191 (2007).
7. Gong, J.， Zou, S.， Liu, D., Li, W., Zhong, Y, & Qiu, X., et al. Reliability and validity of social responsiveness scale. *Medical Innovation of China,* **16**(01), 019-023 (2019).
8. Gross, J. J., & Levenson, R. W. Emotion elicitation using films. *Cognition and Emotion,* **9**(1), 87-108 (1995).
9. Krug, D. A., Arick, J., & Almond, P. Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *Journal of Child Psychology Psychiatry,* **21**(3), 221-229 (1980).
10. Li, J., Zhong, J., Cai, L., Chen, Y., Zhou, M. Comparison of clinical application of three autism rating scales. *Chinese Journal of Contemporary Pediatrics,* **7**(1), 59-62 (2005).