Cognitive Wearable Robotics for Autism Perception Enhancement

MIN CHEN, WENJING XIAO, LONG HU, and YUJUN MA, School of Computer Science and Technology, Huazhong University of Science and Technology, China
YIN ZHANG, School of Information and Communication Engineering, University of Electronic Science and Technology of China, China
GUANGMING TAO, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, China

Autism spectrum disorder (ASD) is a serious hazard to the physical and mental health of children, which limits the social activities of patients throughout their lives and places a heavy burden on families and society. The developments of communication techniques and artificial intelligence (AI) have provided new potential methods for the treatment of autism. The existing treatment systems based on AI for children with ASD focus on detecting health status and developing social skills. However, the contradiction between the terminal interaction capability and availability cannot meet the needs for real application scenarios. At the same time, the lack of diverse data cannot provide individualized care for autistic children. To explore this robot-based approach, a novel AI-based first-view-robot architecture is proposed in this article. By providing care from the first-person perspective, the proposed wearable robot overcomes the difficulty of the absence of cognitive ability in the third-view of traditional robotics and improves the social interaction ability of children with ASD. The first-view-robot architecture meets the requirements of dynamic, individualized, and highly immersed interaction services for autistic children. First, the multi-modal and multi-scene data collection processes of standard, static, and dynamic datasets are introduced in detail. Then, to comprehensively evaluate the learning ability of children with ASD through mental states and external performances, a learning assessment model with emotion correction is proposed. Besides, a wearable robot-assisted environment perception and expression enhancement mechanism for children with ASD is realized by reinforcement learning, which can be adapted to interactive environments with optimal action policies. An interactive testbed for children with ASD treatments is demonstrated and experimental cases for test subjects are presented. Last, three open issues are discussed from data processing, robot designing, and service responding perspectives.

CCS Concepts: • **Human-centered computing → User interface programming**; • **Computer systems organization → External interfaces for robotics;**

This project was supported by the National Key R&D Program of China under grant 2018YFC1314600, the Nature Science Foundation of China under Grant 61802138, 61802139. Prof. Yin Zhang's research is supported by the National Key R&D Program of China (No. 2020YFB1006002). This research is also supported by the Technology Innovation Project of Hubei Province of China under grant 2019AHB061.

Authors' addresses: M. Chen, W. Xiao, L. Hu (corresponding author), and Y. Ma, School of Computer Science and Technology, Huazhong University of Science and Technology, China; emails: {minchen2012, wenjingx}@hust.edu.cn, yujun.hust@gmail.com, hulong@hust.edu.cn; Y. Zhang (corresponding author), School of Information and Communication Engineering, University of Electronic Science and Technology of China, China; email: yin.zhang.cn@ieee.org; G. Tao (corresponding author), Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, China; email: tao@hust.edu.cn.

1533-5399/2021/07-ART97 $15.00
https://doi.org/10.1145/3450630

Additional Key Words and Phrases: Autism therapy, multi-modal and multi-scene, emotion perception, reinforcement learning

ACM Reference format:
Min Chen, Wenjing Xiao, Long Hu, Yujun Ma, Yin Zhang, and Guangming Tao. 2021. Cognitive Wearable Robotics for Autism Perception Enhancement. *ACM Trans. Internet Technol.* 21, 4, Article 97 (July 2021), 16 pages.
https://doi.org/10.1145/3450630

1 INTRODUCTION

Autism spectrum disorder is a general term with a group of neurodevelopmental disorders with the common characteristic of a deviation from normality in manifestations such as social interactions, verbal communication, interests, and behaviors [1]. The latest statistic for the morbidity rate of ASD released by U.S. **Centers for Disease Control and Prevention (CDC)** in 2018 was 1:59 [2], which is 15% higher than the rate of 1:68 that was released in 2016. ASD is seriously endangering the physical and mental health of children. The influences of ASD have not only limited the social activities of patients, such as in social interaction, learning, and working, but they have also placed a heavy burden on families and society. The etiology of autism is still an unsolved problem worldwide, and the existing treatments for autism mainly focus on education and clinical training. Targeted education and training for autistic children younger than 6 years (that is the optimal intervention period) are very likely to produce positive effects on these children. The most typical method is applied behavior analysis proposed by Lovaas in 1987 [3].

However, mandatory training conducted by educators alone is not likely to achieve the expected results and would even cause psychological harm to children with ASD once again. Thus, the traditional treatment methods for autism have numerous limitations. The developments of wireless network communication, **artificial intelligence (AI)**, wearable technology [4], and cloud computing have provided new potential methods for the treatment of autism. There are some researches that have been conducted to develop new computer-assisted methods to improve the current situation in autism treatments [5, 6]. The strengths of the current intervention studies on ASD are as follows: First, compared to traditional manual extraction patterns, the automatic cognitions of information reduce extensive workloads and mitigate the infiuences of subjective factors on results [7]. Second, the automatic feedbacks of interaction alleviate the psychological burdens of autistic children brought by mandatory education [8].

However, there are still two limitations of existing ASD treatment systems based on AI:

- The contradiction between terminal interaction capability and availability: In the current ASD treatment system, the terminal's interaction ability and availability are mutually restricted. There is a system with a relatively robust interaction mechanism to improve the social ability of children with ASD but there are mobility restrictions. Zhao (2018) [9] designed a collaborative virtual reality game to improve the social ability of children with ASD in which a scene is set up to make children with ASD and those without ASD join the game together. However, the game scene has caused an inconvenience for children with ASD and it is not feasible in the actual scene of ASD treatment. The system can meet the requirements of real-time monitoring of the state of children with ASD, but the interaction ability of the terminals is usually insufficient. In Rudovic's 2018 work [10, 11], a personalized machine learning framework was proposed to automatically perceive the emotional state and engagement of children in a robot-assisted autism treatment method. The high mobility of the robot can satisfy the practical requirements of ASD treatment, although the robot's

interaction ability is weak, and it provides only a few simple body movements that are used for interaction with ASD children.

- The lack of diversity data: The data generated by children and related to autism spectrum conditions [12] are really small. It is difficult to find children with ASD due to the small population size. Second, ASD children are always reluctant to interact in a compulsory environment to complete data collection as a result of the nature of the mental illness itself. In addition, guardians can be unwilling to provide data on their children to the disease center due to concerns about privacy protection. The current systems are based on a centralized database for the diagnosis and treatment of children with ASD. The utilization of a deep **convolutional neural network (CNN)** and the ABIDE dataset for the recognition of ASD is discussed [13]. This study also shows the high performance of CNN relative to **support vector machine (SVM)** and random forest. In RadDeep2018 [14], detecting representative activity in the atypical activity of children with ASD through deep learning was proposed, which may help therapists evaluate the effects of behavioural interventions. A centralized database was adopted to conduct one-size-fits-all modeling for all the users in these systems; however, the lack of diversity data cannot provide individualized care for children with ASD.

These existing treatment systems are based on AI methods and aim to protect and improve the social ability of children with ASD [15]. Nevertheless, the contradiction between terminal interaction capability and availability cannot meet the requirements for real application scenarios. At the same time, the lack of diversity data does not provide individualized care for ASD children. To address the challenges faced by the existing treatment systems, a novel type of terminal interaction architecture based on wearable robots is proposed to meet the health monitoring and socialization needs of ASD children in real application scenarios. The proposed wearable robot does not a robot's exoskeleton but is a social-emotional robot with a first-view that can realize emotion perception and interaction [16]. Specifically, the following three aspects of the first-view-robot architecture are considered:

- A dynamic dataset supplement: In the treatment system for children with ASD, the dynamic provision of the dataset is the basis for ensuring the reliability and validity of the system. Individualized treatment is possible with such a dynamic dataset collection process. The dynamic dataset supplement requires the cooperation of the cloud and multiple wearable intelligent terminals. Wireless network technology provides support for high-load data transmission and communication.
- Individualized modeling in the cloud: The cloud provides a platform with rapid computation and storage capabilities for the individualized modeling of children with ASD. A single terminal can learn and analyze in the cloud to recognize the user's state by collecting the terminal data. Furthermore, the sharing of terminal data in the cloud statistically enhances the learning ability of the model. Thus, the requirements of the analysis model are met by high reliability and robustness.
- Immersive interaction with the robot: The social interaction of children with ASD is a cumulative learning process. A robot and an autistic child are a community in which the robot supplements the cognition ability of the child and also guides the behavioral expression of the child with the surrounding environment in the meantime.

In this article, data collection at terminals, model learning in the cloud, and the robot-assisted environment interaction for the proposed treatment system of children with ASD are presented. Specifically, the main contributions of this article are as follows:

(1) An AI-based first-view cognitive wearable robotics architecture for children with ASD is proposed. The proposed architecture considers the heterogeneity of the data, and the collection mechanisms of standard, static, and dynamic datasets are discussed in detail.
(2) A deep architecture-based learning assessment model for children with ASD is developed. The model considers three types of manifestations, including facial expression, body movement, and verbal communication. In addition, the learning assessment results are revised by the emotional state.
(3) A robot-assisted interaction mechanism between children with ASD and the interaction environment is proposed. The environmental perception and the expression enhancement of autistic children are discussed from the perspective of reinforcement learning theory.
(4) A testbed developed for autism treatment is demonstrated, and the test scenes and a test subject case are presented.

The remainder of this article is organized as follows: Section 2 introduces an AI-based first-view cognitive wearable robotics architecture for the healthcare social interaction of children with ASD and describes the different layers involved in the architecture. Section 3 presents the multi-modal data collection mechanisms and feature extraction processes. The learning assessment model with emotion correction for children with ASD is described in Section 4. Section 5 elaborates on the wearable robot-assisted environment perception and expression enhancement mechanism. The system testbed and an experimental case are presented in Section 6 along with a discussion of three open issues for future research. Last, Section 7 summarizes this article.

2 THE SYSTEM ARCHITECTURE

The AI-based first-view paradigm is a new generation of treatment systems for children with ASD using the support of cloud computing and wireless network communication technologies. AI technology provides support for the emotional healthcare of children with ASD. In the proposed first-view-robot architecture, two aspects are considered. The first aspect is that children with ASD lack knowledge of the surrounding social-emotional world [17]. The system needs to provide cognitive assistance to compensate for the lack of cognitive ability of the children with ASD. Furthermore, children with ASD also lack the ability to express their psychological states and physiological behaviors. Here, we put forward the concept of the first-view wearable robot for the first time. The robot refers to all the robots that are sufficiently close to the user and capable of multiple sensor fusion, environmental awareness, and can assist the user to complete cognitive tasks. Compared with the third-view robots in Figure 1, the first-view robots show major advantages: close companion, small volume and light weight, intelligent upgrading, and individualized service and they look at the world and perceive the ambient environment through the user's perspective and finally get more accurate interactive strategies and education programs.

For ASD children, the first-view wearable robot can provide them with the ability to perceive and express the surrounding interactive environment [18], which means that robots and these children stand in the same perspective to perceive the environment and understand emotions. The robot and the ASD child are a community. Moreover, the system can guide to assist in their expression. To provide the learning ability for the ASD child and guide them to conduct interaction with their environment, the robot cognizes the ASD child's state and the surrounding environment. In addition, the wearable robot has the function of a conventional robot, which is called third-view. Third-view care is the general case for the robot. The robot provides diversified interaction, such as voices, stroking, and music broadcasting by recognizing the mental states and the physical manifestations of the ASD children. The proposed first-view-robot architecture is divided into four layers: data collection, communication, cloud analysis, and

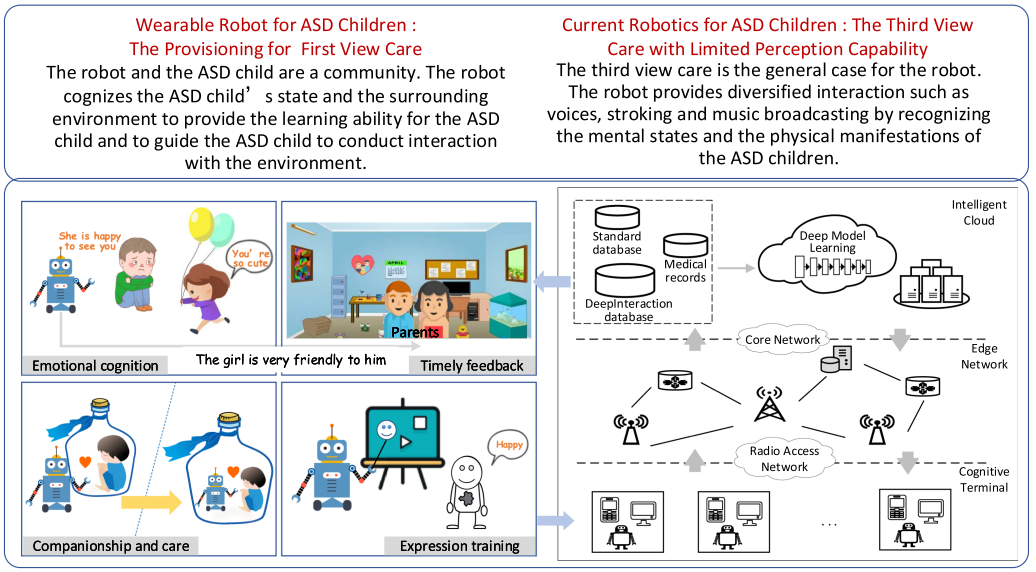


Fig. 1. The AI-based cognitive wearable robotics architecture for enhancing autism perception.

interaction layers, as shown in Figure 1. The design details of these four layers are discussed below.

2.1 The Data Collection Layer

The data collection layer refers to the process of collecting user data by the terminal devices. The terminal devices include a smartphone, microphone, camera, physiological signal sensing device, and the wearable robot. Among these terminal devices, smartphones can collect natural environment information, such as geographical location and weather. The microphone is used to obtain the user's audio signal while the camera is used to capture the user's face image and behavior video. The physiological signal devices are used to perceive the user's physiological signals, including ECG, EEG, and body temperature. Generally, the wearable robot has human behavior patterns, such as voice, walking, stretching arms, and nodding, to interact with users. In addition, the wearable robot can also help users perceive and guide them to express by collecting the social environment information around the user.

2.2 The Communication Layer

The communication layer refers to the data transmission process between the user and the remote cloud platform. The user data collected at the terminal is transmitted to the edge network through the radio access network. The node devices in the edge network complete the data transfer process while providing storage and computing resources for lightweight data processing services. The edge network transmits the user data to the remote cloud platform through the core network. The cloud platform analyzes the user data with computation-intensive service and transmits the analysis results to the user terminal through the edge network.

2.3 The Cloud Analysis Layer

It is difficult to establish an analysis model at the terminal with high reliability and robustness, because the data generated by one patient related to the autism spectrum conditions is small in

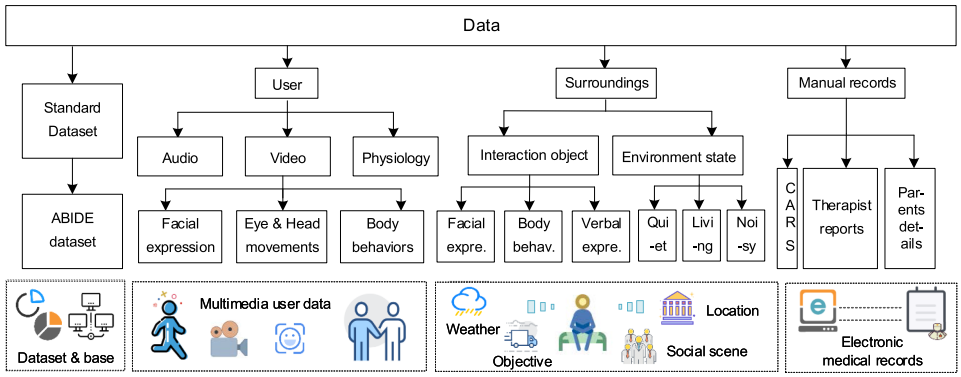


Fig. 2. Composition of the dataset.

size. The cloud analysis layer refers to the data analysis process that is conducted by the data center with powerful computing and storage capabilities. The cloud analysis layer includes two aspects. The cloud collects user data from different terminals and conducts proactive updating for the database and learning model. The cloud databases also include the databases for storing the standard and static datasets. The cloud feeds back the learned model to the user terminal once the model learning is completed. The terminal can quickly construct models and perform real-time analysis after collecting user data.

2.4 The Interaction Layer

The interaction layer refers to the interaction process between the terminal and the user. The cognitive wearable robot guides the user to conduct interaction with the surrounding environment using its body movements and stimulation sources. Meanwhile, the other intelligent terminal devices (audio and video devices) are activated to adjust the psychological states of the user. The wearable robot adjusts its interaction training according to the analysis results fed back by the cloud. The robot-assisted environment perception and expression enhancement enable the user to better understand the intentions of the interactive objects and enhance the user's ability to interact with the environment.

3 MULTI-MODAL DATA COLLECTION AND FEATURE EXTRACTION

3.1 Multi-modal Data Collection

The multimodal data collection includes standard, static, and dynamic datasets. The composition of the dataset is shown in Figure 2. The learning state of children with ASD is related to their mental state, physical state, and behavior performance with the additional influence of the surrounding environment. In this article, multi-dimension data collected from multiple sources is considered to support the learning process of the model with high reliability in the cloud.

3.1.1 Standard Dataset. The autism brain imaging data exchange (ABIDE[1]) database integrates the image data of brain structure and function from multiple laboratories worldwide to accelerate the understanding of the neural mechanism of the brain for autism [19]. The two large data subsets (ABIDE I and ABIDE II) were formed in ABIDE project. The data in ABIDE I and ABIDE II were collected from 1,112 examinees (539 autism patients and 573 normal persons) and 1,114 examinees (521 autism patients and 593 normal persons) in 17 and 19 centers,

[1]ABIDE : http://fcon_1000.projects.nitrc.org/indi/abide/.

respectively. The image data includes sMRI and r-fMRI modes while the other data includes demographic information.

3.1.2 Static Dataset. The static dataset refers to the dataset from medical sites, therapists, and guardians. The **childhood autism rating scale (CARS)** [20] includes 15 items: visual, listening, and taste-smell-touch responses; relationship to people; imitation; body and object use; adaptation to change and use; fear and nervousness; verbal and nonverbal communications; activity level; consistency of intellectual response; and general impressions. The doctors and the guardians jointly grade each item with the relevant behaviors of a child and conduct an evaluation as per the total score. Furthermore, the static dataset also includes periodical analysis reports of the therapist along with detailed records of the guardians.

3.1.3 Dynamic Dataset.

- Audio-based: The audio-based data reflects the verbal communication ability of children with ASD. The voice data of the user is collected through microphones. A time window mode is adopted to trim audio files during the pre-treatment process to ensure the validity of the data. Meanwhile, the beginning and end of an audio file are filtered out. The features of the audio file are extracted after the pre-treatment and the audio feature vectors are obtained.
- Video-based: A video includes multiple effective data that reflect the behavior characteristics of children with ASD. Specifically, the video data includes face images and the repetitive behaviors of the user. Devices, such as cameras, are adopted to collect the video data of the user, and a time window model is used to frame the video files. To obtain the feature vectors of face images, eye movement, head movement, and facial expression features are extracted. Additionally, the feature vectors for repetitive behavior are obtained by analyzing the body movement of the user.
- Physiological signal-based: Additional intuitive data are collected from physiological signals. The physiological signals can reflect the body conditions of children with ASD. Therefore, wearable devices are adopted to collect the physiological signals of the user, including ECG, EEG, body temperature, pulse rate, and breathing. The features of different physiological signals are extracted and the feature vectors for the physiological signals of the user are obtained.
- Environment information-based: The assessment on the learning state of a child with ASD not only requires his or her own data but also data from the surrounding environment with which he or she interacts. Specifically, when a user is in a different interactive environment, such as a study scene, rest scene, entertainment scene, and conversational scene, the environment feature vectors are obtained by collecting the surrounding information of the user.

3.2 Multi-scene Behavior Analysis

The evaluation of the interaction with the surrounding environment of children with ASD is dealt with in different scenes. In the case of a working scene, the communication and the work execution abilities of children with ASD need to be trained. The physical coordination ability and teamwork skills of autistic children need to be analyzed in the case of sports scenes. Therefore, four scenes were considered in this study to evaluate the learning ability of children with ASD: study, rest, entertainment, and dialogue scenes. The different scenes correspond to different time slots in a day to comprehensively conduct the evaluation and enhancement of the learning ability of children with ASD. The different scenes can be distinguished by recognizing the background targets [21]. There are representative background targets in different scenes, such as a computer, running machine, projector, or sofa. The real-time data of children with ASD are collected in a specific scene. The

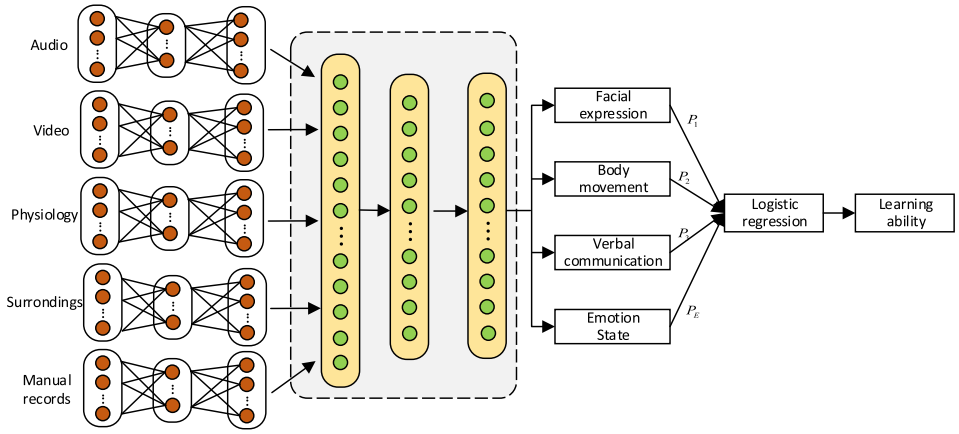


Fig. 3. The deep architecture of learning evaluation for children with ASD.

learning evaluation and optimal environment expression enhancement strategy are provided for that specific scene.

4 LEARNING EVALUATION MODEL WITH EMOTION CORRECTION

The multi-modal data from multiple sources make it possible to construct an efficient model in the cloud. Meanwhile, both the data transmission process at multiple terminals by the network and the cumulative learning process of the model in the cloud jointly ensure the reliable evaluation of the learning state for children with ASD. The deep architecture of the learning evaluation method for children with ASD proposed in this article is shown in Figure 3. The proposed architecture includes the feature extraction, feature fusion, and deep learning process of multi-modal data. Moreover, the emotional state is adopted to correct the results of the learning evaluation.

Since there is a large difference in the learning states of children with ASD at different moments of a day, such as working, resting, entertainment, and sports, the learning state of a child with ASD is evaluated in the specific scene that he or she is in. The proposed architecture collects multi-modal data, and the audio-, video-, physiological signal-, environment information-, and manual records-based feature vectors are obtained as x_a, x_v, x_p, x_s, and x_m, respectively. The feature fusion for different features is conducted and the feature vector set is obtained and recorded as $X = \{x_a, x_v, x_p, x_s, x_m\}$. In-depth features related to the learning ability of children with ASD are deduced through the deep learning of multiple iterative layers. It is assumed that the facial expression, body movement, and verbal communication abilities of children with ASD are recorded as A_1, A_2, and A_3, respectively. The positive and negative samples for the three types of performances are classified using the labeling by the therapists, and the label sets are $\{a_{11}, a_{12}\}$, $\{a_{21}, a_{22}\}$, and $\{a_{31}, a_{32}\}$. The classification probability for these three types of performances are P_1, P_2, and P_3. In consideration of the direct correlation between the mental states of children with ASD and their external manifestations [22], emotion is considered to be a psychological feature in the proposed architecture for conducting the correction of the learning ability of children with ASD. It is assumed that the emotion state set of children with ASD is recorded as $E = \{e_1, e_2, \ldots, e_n, \ldots, e_N\}, (1 \leq n \leq N)$, where N denotes the total number of elements in the emotion set. The classification probability for the emotion state of children with ASD is recorded as P_E. The multivariate logistic regression is adopted for modeling, and the emotion correction evaluation result for learning the state of the children with ASD is as

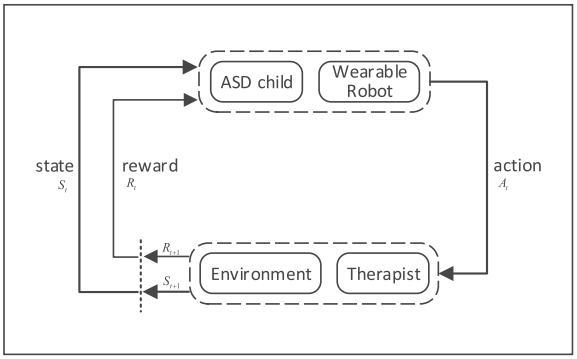


Fig. 4. Learning enhancement mechanism: The diagram of reinforcement learning model for environment-robot-ASD child.

follows:

$$P = w_0 P_E + w_1 P_1 + w_2 P_2 + w_3 P_3, \tag{1}$$

where w_0, w_1, w_2, and w_3 represent the weights allocated for the emotional state, facial expression, body movement, and verbal communication abilities, respectively.

5 THE WEARABLE ROBOT-ASSISTED ENVIRONMENT PERCEPTION AND EXPRESSION ENHANCEMENT MECHANISM

The wearable robot-assisted environment perception and expression enhancement mechanism that is proposed in this article includes two aspects. The first aspect, called the third-person perspective, is the general interaction case between the wearable robot and the children with ASD. The wearable robot conducts emotion-aware interaction and communication with children. The second aspect is the companionship and care by the wearable robot toward the children with ASD from the perspective of the first person. The robot and the child with ASD are a community. The robot provides environmental perception and expression ability for the child through the support of the cloud. The first- and third-person perspectives of the wearable robot are shown in Figure 1.

The third-person perspective is the conventional case for the wearable robot. The wearable robot provides diversified interaction, such as voices, stroking, and music broadcasting by recognizing the mental states and the physical manifestations of children with ASD. Under the first-person perspective, the wearable robot and the ASD child are a community, and thus the learning ability of the ASD child is trained. The ASD children's state and the surrounding environment are cognized by the wearable robot to provide the learning ability for the ASD child and to guide the ASD child in conducting interaction with their environment. Furthermore, the wearable robot feeds back the mental state and the manifestations in interaction with the environment of an ASD child to his or her parents/guardians to enable them to better know and communicate with their child.

In this article, the wearable robot-assisted learning enhancement mechanism for children with ASD is analyzed from the theoretical perspective of reinforcement learning [23]. The wearable cognitive robot-assisted environment perception and expression enhancement mechanism is discussed in the entire learning evaluation system for children with ASD, as shown in Figure 4. The learning state, the surrounding environment, and the guiding action of the wearable robot for an ASD child are assumed to be recorded as S, I, and A, respectively. $Q(S_t, A_t)$ represents the estimation for the total return obtained if action A_t is executed under state S_t with an immediate return R_t. The system aims to provide the optimal behavioral expression in the entire interaction process

Fig. 5. Dataset for training the ASD evaluation model: an example.

between a child with ASD and their environment. When an ASD child is in state S_t, a value of A_t makes the $Q(S_t, A_t)$ maximum is selected. Then, the user state is changed from S_t into S_{t+1}, and the updated value of Q at next moment is as follows:

$$Q(S_t, A_t) \leftarrow (1-\alpha) \cdot Q(S_t, A_t) + \alpha \cdot (R_t + \gamma \cdot \max_A Q(S_{t+1}, A)), \tag{2}$$

where α is the learning rate, γ is the discount factor, and $0 \leq \gamma < 1$. The values of Q and R are jointly provided by the environment and the therapists.

During the interaction process, the autism child and the wearable robot are a community for learning. The wearable robot conducts a learning process of mapping from the environment and the ASD child's state to action to endow the child with an optimal behavior expression. In the meantime, to complete the guidance process of behavior expression for the ASD child, the wearable robot adopts the optimal action strategy with the maximum return value given by the environment and the therapists.

6 EXPERIMENTS AND TESTBED

The AI-based first-view cognitive wearable robotics architecture provide interactive care from the first-person perspective. To prove the effectiveness of the system, a testbed platform for evaluation of autistic children is established. First, a dataset of autism is established, vision-based ASD evaluation model is presented for end-to-end autism detection, which uses three-dimensional face changes as detection features. Then, we build an available testbed for autistic children treatments and carried out the experimental case test for test subjects.

6.1 ASD Evaluation Model

Through the camera of the AIWAC robot, we can collect users' facial data. The facial information is used as features to evaluate their current state on autism. And the appropriate treatment mechanism is selected based on the autism risk assessment results. This section mainly introduces the training details of the ASD evaluation model based on facial features.

A dataset used to train ASD evaluation model is built. We cooperated with mental health medical institutions to conduct cognitive training test for autism. The test subjects are children from volunteer families. Specifically, cognitive training test includes three stages: building blocks game, intelligence testing, and speaking through pictures. The camera in the experimental environment records the performance of children throughout the entire testing process. Based on children's performance in tests and clinical medical analysis from professional doctors, these psychologists will diagnose whether the user is autistic, which will be the label for the data samples. Figure 5 shows an example of dataset for training the ASD evaluation model. There were 124 test subjects in the dataset, including 36 autistic children and 88 normal children. Finally, we obtained video resources and diagnosed results of these participants.

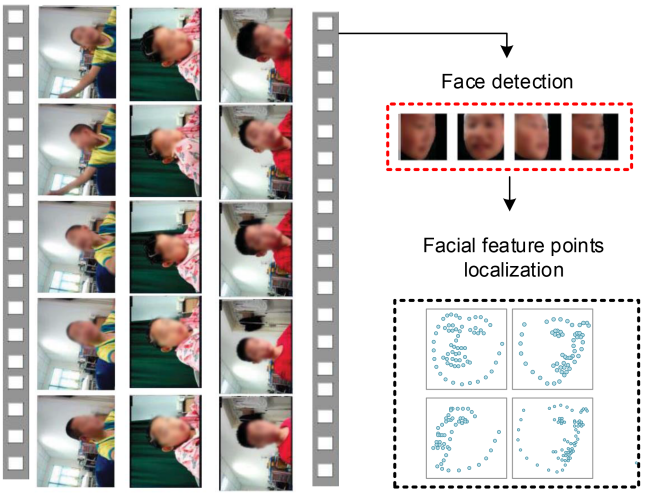


Fig. 6. Facial feature extraction of ASD.

First, OpenCV [24] is used to process the video stream and extract it as frame images with a frame rate of 1 s. Figure 6 shows the process flow of face feature extraction. OpenFace API [25] is a tool for detecting the face area and extracting the key features of the face, and it is a free and open source face recognition project with deep neural networks, and the face detection algorithm filters out background interference and extracts the face features. In our experiment, the selected face feature vector is the face feature position in three-dimensional space. Besides, the euclidean three-dimensional distance and angle value of human face also are add into the feature vector of ASD evaluation model, which expose information about mouth shape features and overall facial expression status. Compared with 2D facial features, facial geometric features and other three-dimensional features can show more specific changes in expressions, making it easier to distinguish different people. We extract a total of 383 face features, and each frame image is converted into a 383-dimensional face feature vector.

In the experiment, we divide the training set and test set according to volunteers, that is, 70% of the user samples are used as the training set, and 30% of the user samples are used as the test set. When capturing video frames, we only selected the clips in which children answered the training questions as valid data and deleted the video clips of children listening to psychologists. Because children are often in a daze, and there is no obvious difference in emotion and facial performance at these times. According to statistics, the effective video clips of users range from 5–12 minutes. And each video data of users will generate multiple data samples, and all data samples of a person only appear completely in the training set or the test set.

For the classifier, a fully connected network is used to establish the ASD evaluation network model, as shown in Figure 7. The input of the evaluation network is the face feature vector discussed above, and the low-dimensional features are mapped to the high-dimensional space through the evaluation network to predict autism risk level. The fully connected network has five hidden layers. The first hidden layer has 1,024 processing units, the second layer has 512 processing units, and the third, fourth and fifth layers are set to 256, 128, and 64 processing units, respectively. To prevent overfitting, a relu function is added after each hidden layer. In addition, we choose two traditional machine learning algorithms as basic algorithm to compare the evaluation effects of our model. The results are shown in Table 1.

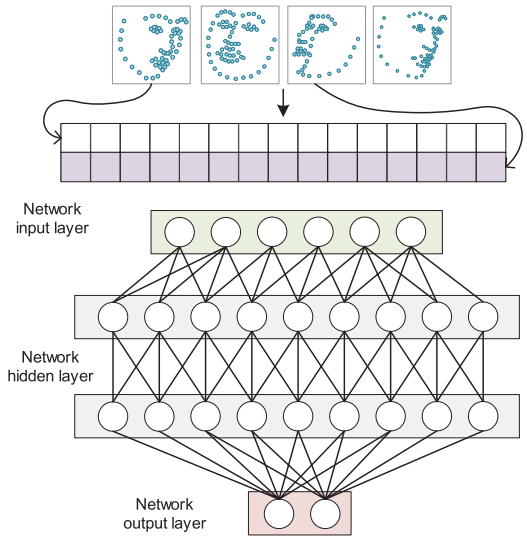


Fig. 7. ASD evaluation network structure.

Table 1. Results of Fully Connected Network and Machine Learning Models

Algorithm	Train Accuracy	Test Accuracy
Full-connect network	88.83%	82.83%
SVM	80.21%	78.47%
XGBoost	83.54%	71.15%

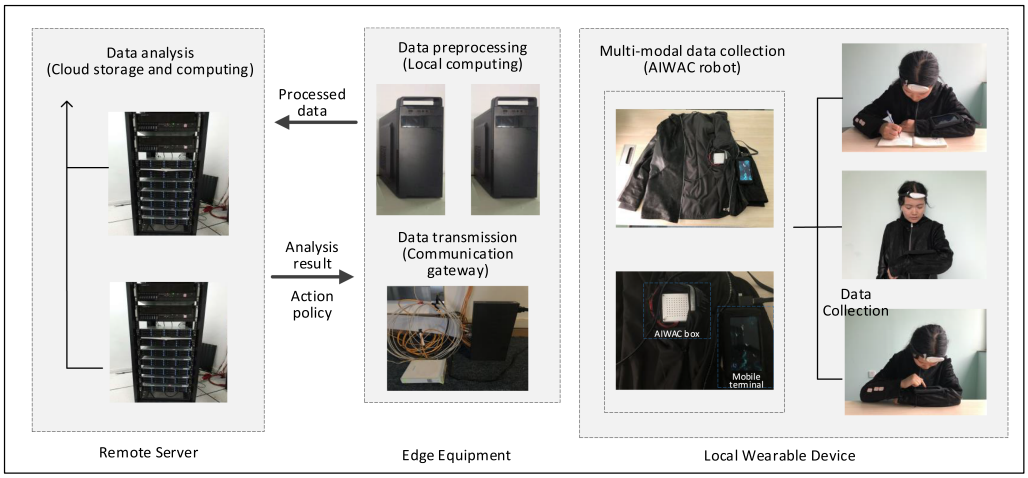
It can be seen in Table 1 that the prediction effect of the fully connected network model is the best. But, we found that using machine learning algorithms also has reliable results. In the actual deployment of the model, different algorithms can be selected according to the resource and performance requirements of the system. For medical institutions, they have sufficient communication and computing equipment and can choose a fully connected network model. In the home environment, it is more convenient to use mobile software to conduct cognitive training for children, and machine learning algorithms will be more suitable.

6.2 Testbed and Experimental Cases

A testbed for autism state evaluation and emotion interaction is introduced in this section. The testbed was developed by the Embedded and Pervasive Computing Laboratory of Huazhong University of Science and Technology. The aim of this testbed is to provide users with autism state detection and emotion interaction services.

The above-mentioned testbed is composed of an AWAIC robot [16], a communication gateway, a local computer, and a remote cloud, as shown in Figure 8(a). The functions and the communication processes among the devices are described below.

- The AWAIC robot, a wearable fashion clothing item, integrates a microphone and an AIWAC box used for voice interaction and a separate mobile terminal that can implement different application scenarios. The user interacts in a specific application scenario with the



(a) Composition of the testbed.

(b) Analysis report at the mobile terminal.

Fig. 8. Testbed and mobile application at user terminal.

AIWAC robot at the same time, which is utilized to analyze the user's psychological and emotional state to enhance the user's interactive experience and interactive performance. The user data are collected not only from devices equipped by the AIWAC robot but also from other intelligent devices cognized by the robot, such as the brain link (the brain-wearable device).

- The communication gateway is the bridge for communication among the local devices and also between the local device and the remote cloud. The data transmissions include transmission from the AWAIC robot to the local computer, the transmission from the local computer to the remote cloud, and the transmission from the remote cloud to the AWAIC robot.
- The local computer provides local computing for data filtering and pre-processing. The pre-processed data not only improves the data quality of the remote cloud database and the accuracy of the analysis model but also reduces the load of network transmission to improve the network performance.
- The remote server provides cloud-based storage and computing services. The pre-processed multi-modal data is stored in the database in the cloud. The cloud-based computing module and cloud database constantly enhance the learning ability and prediction accuracy of the

model. Using the communication module of the cloud, the analysis results along with the action policies are fed back to the robot.

- The terminal receives feedback from the cloud and generates evaluation reports to monitor the conditions of the user. Meanwhile, to facilitate better communication between the guardians and the user, visual interfaces are provided for browsing and viewing by the guardians.

The experimental process and details are described (see Figure 6(a)). The external microphone of the wearable robot was used to record voices. Meanwhile, the brain-link worn by the test subject was used to collect the EEG signals. The experimental process was recorded as a video file by the camera.

The evaluation report of the test subject was generated in the ChildrenCare App that was deployed at the mobile phone terminal. Statistical evaluation was conducted regarding three aspects: autism evaluation results, emotional state, and activity statistics. And autism evaluation results are obtained by the autism assessment model analysis (in Section 6.1). The emotional state and activity statistics were from AIWAC robot analysis module. The statistics of a test subject that was conducted by the system over three consecutive days are shown in Figure 6(b). First, the percentage for the autism evaluation result of the test subject is provided in the report. The larger the percentage number is, the higher the degree of autism. It can be seen from the figure that the autism state of the test subject was steady at about 60 percent. In the second part, the emotions of the test subject throughout a day are recorded by the Mood Chart. The different moods monitored include happiness, calmness, depression, sadness, and anger. As shown, the mood of the test subject early in the day was relatively calm. However, the test subject's mood at the end of the day was extreme, and anger occurred. In the third part, the statistics were conducted by the Activity Count for the number of different activities (including five different types of activities: learning, rest, work, conversation, and entertainment) of the test subject throughout a day. As shown, learning, rest, and entertainment accounted for the majority of the daily activities of the test subject.

7 OPEN ISSUES AND CONCLUSION

7.1 Open Issues and Future Directions

To improve the accuracy in autism evaluation for users, multi-modal data were collected through multiple intelligent terminals that provide support for effective deep learning of the cloud model. The reliability and robustness of the model were enhanced with wide data collection and deep model learning. It is difficult to evaluate which kind of data is effective for the analysis of autism. The autism model should be established based on factors that influence the autism state of the user. Furthermore, the transmission of effective data can reduce the network load and improve network performance [26]. Therefore, an effective data processing method should be designed.

The wearable robot and the ASD child are a community, which further enhances the appetency of the wearable robot. Moreover, the hidden dangers in anthropogenic training and interaction are reduced. The geographical location of the user may often change, and it is inconvenient to carry around a household health robot. The authors are planning to develop a wearable affective robot. A convenient and effective terminal for the interaction robot will be designed that integrates the wearable clothes concept with the robot mechanism. The robot will integrate multiple sensors to collect user data in real time and to provide monitoring services. Therefore, a new robot designing method is required, which is another open and challenging problem to be addressed in the future.

The cloud acts as the smart brain of the wearable robot and constantly stores multi-modal data to enhance the learning ability of the model and to provide behavior-guiding strategy. It is essential that high-reliability and low-latency service response mechanisms are provided when the

interaction scenarios and the objects are changed. Furthermore, the strategy for the wearable robot to guide the interaction under different scenes should be given by the cloud to provide services that meet user demand and ensure the quality of experience. Therefore, another future direction for research consideration is in the development of an effective network management method and elaboration of interaction strategies.

7.2 Conclusion

In this article, a highly immersive, dynamic, and individualized treatment system for children with ASD is proposed. The proposed treatment system evaluates the learning state of children with ASD and strengthens their environmental perception and expression through the dynamic collection of multi-modal data, individualized model learning in the cloud, and a wearable robot-assisted interaction in a closed-loop framework. An AI-based first-view-robot architecture for children with ASD is presented. The standard dataset, static dataset, and dynamic data collecting process from multiple terminals were discussed, and behavior analysis for children with ASD was discussed in various scenes. A learning assessment model for children with ASD with emotion correction was provided that can comprehensively evaluate the user emotions, facial expression ability, body movement ability, and language expression ability. Furthermore, a wearable robot-assisted environment perception and expression enhancement mechanism for children with ASD was discussed from the theoretical perspective of reinforcement learning. The learning process of the robot for the surrounding environment is mapped to the optimal behavior guiding process for children with ASD. A testbed for autism treatment was established. The test subject conducted activities in four different scenes: learning, rest, entertainment, and dialogue scenes. The evaluation report for the test subject consisted of evaluation results for autism state, emotion chart, and statistics for the number of activities was provided by the app in a mobile phone terminal. Finally, three open problems and future research directions were discussed, which consist of finding an effective data processing method, enhancing wearable robot-based emotion interaction, and achieving service and experience quality assurance with high reliability and low latency.

REFERENCES

[1] L. Fusar-Poli, N. Brondino, P. Politi, and E. Aguglia. 2020. Prevalence of autism spectrum disorder among children aged 8 years - Autism and developmental disabilities monitoring network. *Eur. Arch. Psych. Clin. Neurosci.* (2020). DOI: 10.1007/s00406-020-01189-w

[2] B.-Y. Park, S. J. Hong, S. L. Valk, et al. 2021. Differences in subcortico-cortical interactions identified from connectome and microcircuit models in autism. *Nat Commun.* 12, 2225 (2021). https://doi.org/10.1038/s41467-021-21732-0

[3] L. Mottron and D. Bzdok. 2020. Autism spectrum heterogeneity: fact or artifact? *Mol Psychiatry* 25 (2020), 3178–3185. https://doi.org/10.1038/s41380-020-0748-y

[4] M. Chen, Y. Jiang, N. Guizani, J. Zhou, G. Tao, J. Yin, and K. Hwang. 2020. Living with I-fabric: Smart living powered by intelligent fabric and deep analytics. *IEEE Netw.* 34, 5 (2020), 156–163.

[5] S. S. Alwakeel, B. Alhalabi, H. Aggoune, et al. 2015. A machine learning based WSN system for autism activity recognition. In *IEEE 14th International Conference on Machine Learning and Applications (ICMLA).* 771–776.

[6] W. Liu, M. Li, and L. Yi. 2016. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Res.* 9, 8 (2016), 888–898.

[7] L. A. Livingston, P. Shah, V. Milner, et al. 2020. Quantifying compensatory strategies in adults with and without diagnosed autism. *Molecular Autism* 11, 15 (2020). https://doi.org/10.1186/s13229-019-0308-y

[8] P. G. Esteban, P. Baxter, T. Belpaeme, et al. 2017. How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn, J. Behav. Robot.* 8, 1 (2017), 18–38.

[9] H. Zhao, A. R. Swanson, A. S. Weitlauf, Z. E. Warren, and Nilanjan Sarkar. 2018. Hand-in-hand: A communication-enhancement collaborative virtual reality system for promoting social interaction in children with autism spectrum disorders. *IEEE Trans. Hum.-mach. Syst.* 48, 2 (2018), 136–148.

[10] O. Rudovic, J. Lee, M. Dai, et al. 2018. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci. Robot.* 3, 19 (2018).

[11] F. Ke, J. Moon, and Z. Sokolikj. 2020. Virtual reality based social skills training for children with autism spectrum disorder. *J. Spec. Educ. Technol.* (2020). DOI: 10.1177/0162643420945603
[12] M. Eni, I. Dinstein, M. Ilan, I. Menashe, G. Meiri, and Y. Zigel. 2020. Estimating autism severity in young children from speech signals using a deep neural network. *IEEE Access* 8 (2020), 139489–139500.
[13] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, et al. 2018. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroIm.: Clin.* 17 (2018), 16–23.
[14] N. M. Rad, S. M. Kia, C. Zarbo, et al. 2018. Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. *Sig. Process.* 144 (2018), 180–191.
[15] F. Thabtah and D. Peebles. 2020. A new machine learning model based on induction of rules for autism detection. *Health Inform. J.* 26, 1 (2020), 264–286.
[16] M. Chen, J. Zhou, G. Tao, et al. 2018. Wearable affective robot. *IEEE Access* 6 (2018), 64766–64776.
[17] C. A. G. J. Huijnen, H. A. M. D. Verreussel-Willen, M. A. S. Lexis, et al. 2021. Robot KASPAR as mediator in making contact with children with Autism: A pilot study. *Int. J. Soc. Robotics* 13 (2021), 237–249. https://doi.org/10.1007/s12369-020-00633-0
[18] C. C. Cheroni, N. Caporale, and G. Testa. 2020. Autism spectrum disorder at the crossroad between genes and environment: Contributions, convergences, and interactions in ASD developmental pathophysiology. *Molec. Aut.* 11, 1 (2020), 69.
[19] A. D. Martino, C. G. Yan, Q. Li, et al. 2014. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molec. Psych.* 19, 6 (2014), 659–667.
[20] K. K. Mujeeb Rahman and M. Monica Subashini. 2021. A Deep Neural Network-Based Model for Screening Autism Spectrum Disorder Using the Quantitative Checklist for Autism in Toddlers (QCHAT). *J. Autism Dev Disord* (2021). https://doi.org/10.1007/s10803-021-05141-2
[21] Y. Liao, S. Kodagoda, Y. Wang, et al. 2016. Understand scene categories by objects: A semantic regularized scene classifier using convolutional neural networks. In *IEEE International Conference on Robotics and Automation (ICRA).* 2318–2325.
[22] B. Weiner. 1980. A cognitive (attribution)-emotion-action model of motivated behavior: An analysis of judgments of help-giving. *J. Person. Soc. Psychol.* 39, 2 (1980), 186–200.
[23] R. S. Sutton and A. G. Barto. 2018. *Reinforcement Learning: An Introduction* (2nd ed.). The MIT Press.
[24] D. I. Katzourakis, E. Velenis, D. Abbink, R. Happee, and E. Holweg. 2012. Race-car instrumentation for driving behavior studies. *IEEE Trans. Instrum. Meas.* 61, 2 (2012), 462–474. DOI: 10.1109/TIM.2011.2164281
[25] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency. 2018. OpenFace 2.0: Facial behavior analysis toolkit. In *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG'18).* 59–66. DOI: 10.1109/FG.2018.00019
[26] M. Chen and Y. Hao. 2020. Label-less learning for emotion cognition. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 7 (2020), 2430–2440.

Received October 2020; revised January 2021; accepted February 2021