



# Interactive Robot-Aided Diagnosis System for Children with Autism Spectrum Disorder

Szu-Yin Lin<sup>1</sup>(✉), Yi-Pei Lai<sup>1</sup>, Hao-Chun Chiang<sup>1</sup>, Yawei Cheng<sup>2</sup>, and Shih-Yi Chien<sup>3</sup>

<sup>1</sup> Department of Computer Science and Information Engineering, National Ilan University, Yilan, Taiwan

szuyin@niu.edu.tw

<sup>2</sup> Institute of Neuroscience and Brain Research Center, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

<sup>3</sup> Department of Management Information System, National Chengchi University, Taipei, Taiwan

**Abstract.** Autism spectrum disorder (ASD) is a group of complex neurodevelopmental disorders characterized by difficulties with social communication and interaction as well as restrictive interest and stereotyped behavior. Despite the behavioral symptoms of ASD often appear early in infancy, the ASD diagnosis is often cumbersome even for expert clinicians owing to characteristic heterogeneity in the symptoms and severity. Early diagnosis and intervention can help children with ASD to achieve more improvement, particularly in their social communication. Here, the study designs an interactive robotic agent and an intelligent image analysis system to assist in the ASD diagnosis of children. The children's facial expression images and body pose movement images are collected during the human-robot interaction, which three computational models are used for further data analysis. The stored database is presented as a reference for diagnosis in a visual interface. Furthermore, we incorporate multiple AI models in facial emotion recognition and eye tracking detection to automatically analyze images and visualize data, assisting clinicians in diagnostic decision making.

**Keywords:** autism spectrum disorder (ASD) · human-robot interaction · robot-aided diagnosis · computer vision

## 1 Introduction

Autism Spectrum Disorder (ASD), based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [1], is a complex group of neurodevelopmental disorders characterized by social communication difficulties, restrictive interests, and repetitive behaviors. In terms of social communication difficulties, patients may show (1) a lack of eye contact, (2) non-verbal communication difficulties, and (3) difficulties in developing, maintaining, and understanding social relationships [1–3]. In terms of repetitive behaviors and restrictive interests, there may be (1) repeated use of specific objects or language, as well as persistent repetition of the same actions, (2) adherence to fixed

rules, fixed speaking patterns, or expressing oneself through specific physical gestures, (3) highly restricted and fixed interests, and (4) high or low sensory responses or strong interests in environmental stimuli [1–3]. Early intervention as possible can help children to make up for inherent defects and fully unleash potential, and also aid in improving social difficulties in ASD. However, clinicians usually take time to diagnose ASD in children with the need of intensive clinical training, catching subtle facial emotions, eye contacts, and gestures during clinical practice. The care-givers or parents are required to fill out a series of psychological assessments. Only after the physicians conduct a comprehensive evaluation can the diagnosis of ASD be confirmed [2].

A simple definition of Human-Robot Interaction (HRI) is the dynamic relationship between humans and intelligent robots [5]. Many studies are dedicated to using robots in treating autism, where robots serve as the interacting partner with the patient to achieve therapeutic outcomes. The use of robots in [6] to treat communication obstacles in patients resulted in significant improvement in communication compared to other patients. The symptoms displayed by individuals with autism are often complex, and many patients have difficulties in communication and interaction. This increases the difficulty in diagnosis, as physicians need to spend a significant amount of time building relationships and emotions with the children to better initiate interaction and observe the child's behavior for diagnostic purposes. As in [7–8], research teams proposed using robots in four test games with children. The children's success in completing the tasks was marked by both the robots and medical professionals. The final markings were compared to each other, and the model of the robot coding the participants' behavior has adjusted accordingly, allowing the robot to automate the interaction tasks and observe the participants' behavior. The information was then provided to physicians as a reference through coding. In the study [9], a parrot-like robot was placed in a room, and the interaction between the robot and the child was recorded for about 190 s. Algorithms were used to calculate the child's position relative to the robot, and features were extracted and classified using a Gaussian support vector machine to determine if the subject was an autistic child. Other studies have also used small tabletop robots for interaction with children. The research used the children's facial features and movement features as the data set and trained the model, allowing the model to recognize children at risk of autism [10].

In the era of rapid development of hardware and artificial intelligence, this study references the groundbreaking work of Sylvia and Ricky Emanuel, pediatricians and psychotherapists at the University of Edinburgh, who used the mobile turtle-shaped robot LOGO for autism therapy in 1976 [4]. Many research labs have also begun to study the possibility of using robots as a therapy for children with autism since the late 1990s. Based on the above research, the goal of this study is to integrate the diagnostic process into the robot and use the robot as a tool to assist physicians in diagnosing children with autism through interactions with the children. The diagnostic process also involves collecting observation information from professional doctors and primary caregivers through a mobile app and collecting related image data through the robot's camera and external cameras. The collected data is analyzed and stored, and the result of the data

analysis is visualized for the doctor as a reference for diagnosing and monitoring the condition. The following lists the main research objectives of this study.

1. Implement an automated human-machine interactive system for evaluating and monitoring autistic children and addressing the challenge of multiple evaluations and monitoring required for autism diagnosis.
2. Gather interaction and image data during human-robot interaction and establish a database that can aid physicians in diagnosis and image analysis.
3. Develop an AI-powered image analysis method for detecting body pose movements, facial expressions, and eye gaze during human-robot interaction to gather information for assessing autism in children and provide it to physicians for diagnosis.

## **2 Interactive Robot-Aided Diagnosis System**

### **2.1 Research Framework**

The framework of the research is shown in Fig. 1 which is divided into four parts: autism diagnosis process and robot interface, observer record, artificial intelligence (AI) extraction of autism children's assessment information, and assisting physicians in diagnosis. In the part of the autism diagnosis process and robot interface, data collection is the main focus. By using Zenbo and its built-in camera, we can interact with children through the robot and enable the camera to record image information in specific processes. The collected image information will be uploaded to a database for storage. The edge AI device NVIDIA Jetson Nano is utilized for detecting the body and specific actions, and the recorded frequency of these actions is uploaded to the database at the end of the process. With regard to human-computer interaction, there will also be a questionnaire for the main caregivers or nursing staff present to fill out, and the results will be stored in the database. In the part of Artificial Intelligence (AI) extraction of autism children's assessment information, facial expression recognition and eye gaze detection are performed through AI models. The data source is the image information collected in the previous part. The emotion analysis model will classify the characters' expressions in the image, calculate the number of occurrences of various emotions in the subjects during each process, and store the number in the database. The eye gaze analysis model classifies the direction of the eyes' gaze and similarly calculates the number of times the direction of the eyes' gaze is seen in each process. The final value is stored in the database. After the analysis results are completed, data visualization techniques will be used to convert the information such as emotions, eye gazes, poses, and questionnaire results from text and numbers into charts and provide them to physicians as reference materials for diagnostic evaluation.

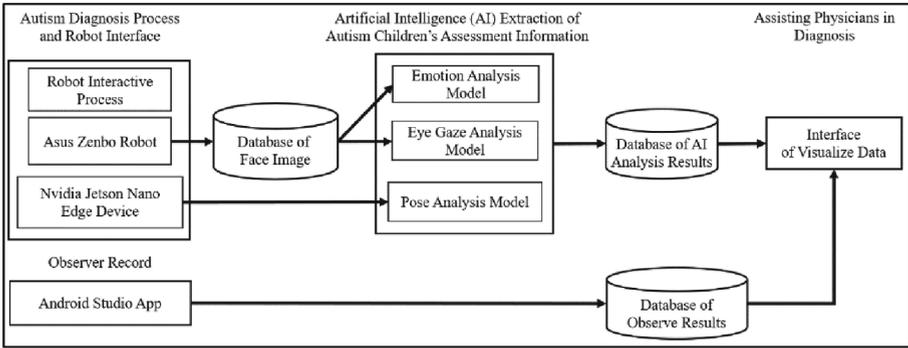
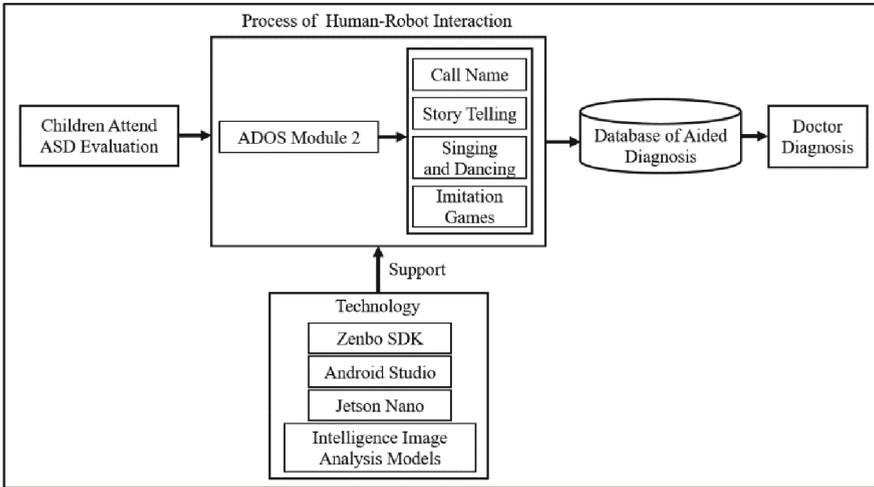


Fig. 1. The research framework of the interactive robot-aided diagnosis system

### 2.2 Autism Diagnosis Process and Robot Interface

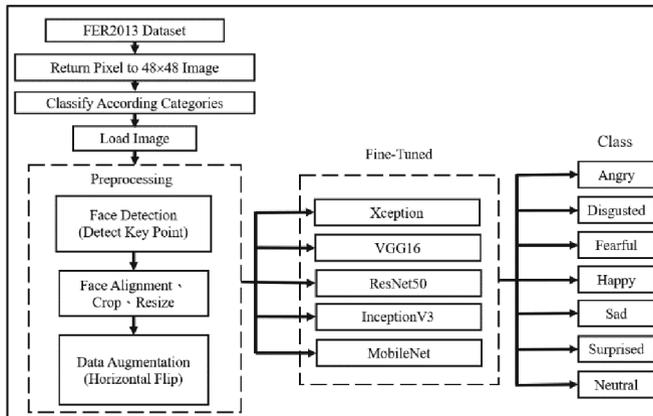
The process of autism diagnosis and human-robot interaction through the robot interface is shown in Fig. 2. This research refers to Module 2 of the Autism Diagnostic Observation Schedule-2nd edition (ADOS-2) in the process. The ADOS-2 (Autism Diagnostic Observation Schedule-2nd edition, ADOS-2) [11] is a clinical assessment tool for physicians, consisting of five modules designed for different age groups. Module 2 is used for interaction and testing with patients with the lowest level of language and communication skills, consisting of 14 activities such as responding to name calls, trust games, telling stories, sharing attention, blowing bubbles, etc. The design of the interactive content is mainly used to observe the patient's communication and social interaction, as well as the presence of any restricted and repetitive behaviors. The checklist also provides key observations in each item of key observation, such as unusual eye contact, facial expressions, shared attention, unusual interests, and rigid behavior, which satisfy clinical physicians' needs to observe Autism during the interaction process [11, 12]. The study designed four human-robot interaction processes: calling names, telling stories, singing and dancing, and playing imitation games. These four processes correspond to the items observed in module 2 of ADOS-2, such as children's facial expression information, eye gaze information, and larger body pose movements during the process. The human-robot interaction process is realized through the support of the Asus Zenbo Robot. The programs needed for the robot process were completed through Android Studio and Zenbo SDK, with the assistance of Jetson Nano for simultaneous intelligent image analysis model operation.



**Fig. 2.** The robot interface and diagnostic process for autism

### 2.3 Using AI to Obtain Assessment Information for Children with Autism

During the evaluation process, the system will record the participant's facial images and body pose movements. Emotions and eye gaze focus will be analyzed based on the facial image information. This experiment aims to conduct experiments with different methods and evaluate the methods that are suitable for the system.



**Fig. 3.** Emotion analysis model training process

**Emotion Analysis Model.** The research uses the FER2013 dataset [13, 14] provided by the Kaggle competition website as training and validation data. FER2013 is a well-known face expression image dataset that is widely known and easy to obtain. Figure 3 shows

the training process of the model used in this research. First, the CSV file of FER2013 is read, which contains the images' pixel values and expression labels. The labels range from 0 to 6 and correspond to angry, disgusted, fearful, happy, sad, surprised, and neutral. Then, the values are restored to  $48 \times 48$  black-and-white images through the Python PIL package. Each image is processed for face alignment, cropping, and enlargement, and a horizontal flip is performed to augment the training data. The training data is input into seven different models for model tuning. The fine-tuned models can output the classification results for the seven expressions.

**Eye Gaze Analysis Model.** The eye gaze analysis model used the Eye-Chimera dataset [15, 16], which consisted of 1,135 image data and was labeled in seven directions: right-up, right, right-down, left-up, left, left-down, and center. In the eye gaze analysis model experiment, three machine learning methods from scikit-learn [17] were compared: Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), and Nearest Centroid. The eye gaze analysis model training process is shown in Fig. 4. First, after the image is read, Dlib's Landmark is used for face detection, and 68 facial key points are marked. The key points of the left and right eyes are 42 to 47 and 36 to 41, respectively. After capturing the human eye key points, the part of the eye is cropped. The image is normalized and standardized to make all images the same size and compress the pixel values between 0 and 1. Finally, the SVM is input to train the model to classify the data into seven categories.

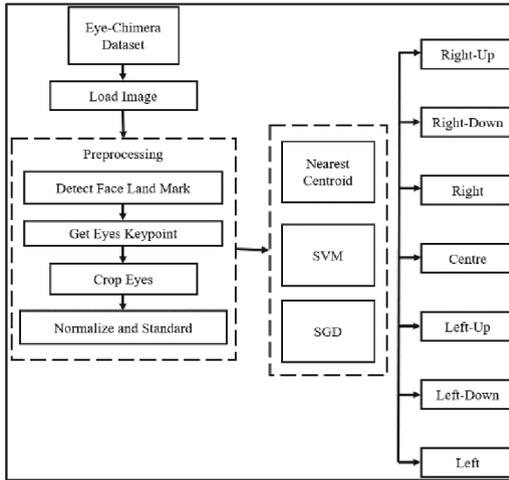


Fig. 4. Eye gaze analysis model training process

**Pose Analysis Model.** The process of the pose analysis model in this study is shown in Fig. 5. The pose analysis model used was implemented by Openpose [18, 19]. The training samples for the joint key points marked by Openpose were mainly from the COCO database, which defined a total of 18 joint parts. After the human body key points are found, the image is judged to have the situation of raising hands, nodding,

and rough movements through the calculation of joint angles and the displacement of facial feature points. The process on the left is the hand-raising detection process. The middle is the detection of nodding and shaking, with the nose as the reference point. When the nose's horizontal or vertical coordinate displacement difference exceeds the set threshold, it is determined that there is nodding or shaking behavior. The process on the right is the detection of rough movements, which uses the elbow, knee, and ankle joints on both sides as reference points. When the joint angle change exceeds the set threshold, it is determined that rough movements have occurred.

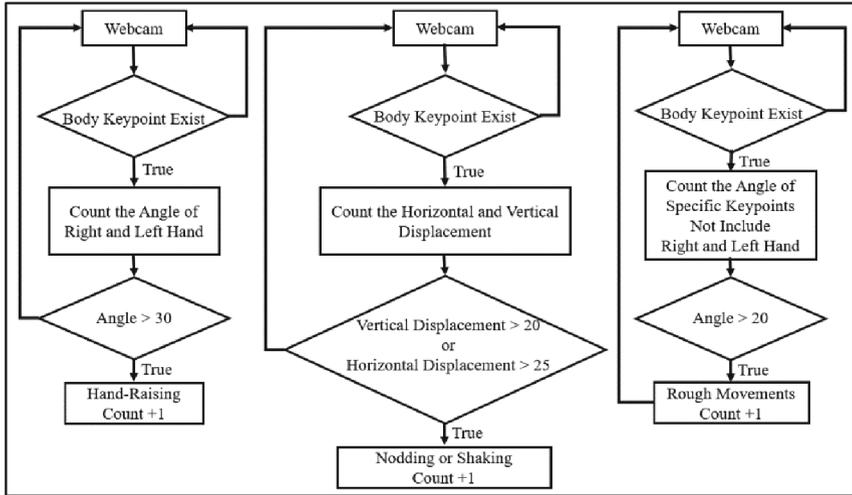


Fig. 5. Three detection processes in the pose analysis model

## 3 Experiments

### 3.1 Emotion Analysis Model

This study used the FER-2013 image dataset for model training and validation to develop an emotion recognition model suitable for the system. The following will be a discussion of the experimental results. This experiment used pre-trained models on Keras, and the weights selected were trained on ImageNet [20]. The models used are VGG16, ResNet50, InceptionV3, Xception, and MobileNet.

**Experimental Results and Discussion.** The results of this experiment were evaluated using Precision, Recall, and F1-score as indicators. Table 1 shows the F1-score of five models, and it can be seen from the table that in terms of the F1-score indicator, InceptionV3 and Xception are relatively better-performing models. By comparing the performance of the two models in each emotion through Precision and Recall in Table 2, we finally decided to use Xception as the model for emotion analysis.

**Table 1.** F1-score of the five models in the seven emotion classification

F1-score	Angry	Disgusted	Fearful	Happy	Sad	Surprised	Neutral	Avg
VGG16	0.55	0.62	0.51	0.87	0.44	0.80	0.63	0.63
ResNet50	0.57	0.68	0.53	0.87	0.43	0.81	0.63	0.65
InceptionV3	0.55	0.63	0.53	0.87	0.48	0.82	0.65	0.65
Xception	0.56	0.69	0.53	0.87	0.46	0.82	0.65	0.65
MobileNet	0.57	0.69	0.48	0.87	0.45	0.80	0.65	0.64

**Table 2.** Precision/Recall of the five models in the seven emotion classification

Precision/ Recall	Angry		Disgusted		Fearful		Happy		Sad		Surprised		Neutral		Avg	
	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re
VGG16	0.51	0.60	0.76	0.53	0.46	0.57	0.90	0.84	0.52	0.39	0.80	0.79	0.63	0.64	0.65	0.62
ResNet50	0.59	0.54	0.92	0.54	0.50	0.57	0.84	0.89	0.56	0.35	0.90	0.83	0.59	0.68	0.70	0.63
InceptionV3	0.60	0.50	0.57	0.71	0.51	0.54	0.86	0.87	0.52	0.45	0.81	0.83	0.61	0.69	0.64	0.66
Xception	0.59	0.54	0.72	0.66	0.52	0.54	0.89	0.86	0.52	0.42	0.82	0.82	0.59	0.73	0.66	0.65
MobileNet	0.53	0.61	0.87	0.56	0.57	0.41	0.88	0.86	0.47	0.44	0.76	0.85	0.61	0.69	0.67	0.63

### 3.2 Eye Gaze Analysis Model

In this study, the Eye-Chimera image dataset was used for model training and testing, and an attempt was made to find an eye gaze analysis model suitable for the system. The machine learning methods used in the experiment are SVM, SGD-Classifier, and NearestCentroid, all from Scikit learn. The following is a discussion of the experiment results.

**Experimental Results and Discussion.** Table 3 shows the results obtained from the experiment using the Eye-Chimera eye movement dataset with SVM. The table shows that the average accuracy of SVM is 0.8, and the performance of Recall is also 0.83 on average. In general, SVM performs better than SGD-Classifier and NearestCentroid. The model is more average and stable in classifying the seven different eye gaze directions.

**Table 3.** Result of eye gaze analysis using SVM

SVM	Centre	UpRight	UpLeft	Right	Left	DownRight	DownLeft	Avg
Precision	1	0.89	0.65	0.81	0.76	0.62	0.89	0.8
Recall	0.89	0.91	0.91	0.87	0.67	0.8	0.73	0.83
F1-score	0.94	0.9	0.76	0.84	0.71	0.7	0.8	0.81

From Table 4, we can see that SGD-Classifier performed particularly poorly in the classification of “Right,” which may be due to the similarity between the images of

“Right,” “Right-Up,” and “Right-Down.” Although the Precision of the SGD-Classifier for “Right-Up” reached 1.00 in the table, its Recall was only 0.64, meaning that some of the actual “Right-Up” images were predicted to be in other directions. Overall, SGD-Classifier only performed well in the “Centre” category, while it performed poorly in other categories, especially in the classification of “Right.”

**Table 4.** Result of eye gaze analysis using SGD-Classifier

SGD-C	Centre	UpRight	UpLeft	Right	Left	DownRight	DownLeft	Avg
Precision	0.91	1	0.72	0.36	0.76	0.81	0.49	0.72
Recall	0.97	0.64	0.87	0.71	0.6	0.57	0.86	0.75
F1-score	0.94	0.78	0.79	0.48	0.67	0.67	0.62	0.71

Table 5 shows that NearestCentroid generally performed poorly in the seven-direction classification tasks. Only the Centre achieved around 80% Precision among the seven directions, while the results in other directions were poorer. This may be due to NearestCentroid’s method, which compares the distance between the center of each category and the new data and categorizes based on the nearest. However, besides the Centre, the other directions are more similar and close, such as Right-Up, Right, and Right-Down, which mostly look right, leading to misclassification. Overall, SVM performed the best among the three experiments, and as a result, we finally chose SVM as the model for eye gaze and pose analysis.

**Table 5.** Result of eye gaze analysis using NearestCentroid

NC	Centre	UpRight	UpLeft	Right	Left	DownRight	DownLeft	Avg
Precision	0.81	0.74	0.65	0.52	0.62	0.66	0.7	0.67
Recall	0.86	0.69	0.83	0.65	0.5	0.62	0.57	0.67
F1-score	0.83	0.71	0.73	0.58	0.55	0.64	0.63	0.67

### 3.3 Pose Analysis Model

In this experiment, Openpose is used as the main tool for skeleton detection. Openpose adopts the Part Affinity Fields for Part Association (PAF) method for human skeleton detection, which is based on finding the relationship between the body parts and the individuals in the image. First, a set of detected body parts is given, and an attempt is made to assemble these points into a full-body posture. Then, calculations are performed on each body part to find the possible main body position and direction. Using the skeleton information detected by Openpose, displacement or angle calculations are performed to

obtain the desired observed actions, such as nodding, shaking, rough movements, and raising hands. In the implementation, Jetson Nano is used as the edge computing device to execute Openpose and the program that calculates the changes in limb displacement or angles.

### 3.4 Summary

The study employed publicly available datasets to evaluate the models used in the system and determine the most appropriate models. To ensure our system diagnoses are reliable and can serve as a robust reference for doctors, a model that can handle a wide range of emotions was selected rather than only processing a small set of features. Xception was therefore chosen as the model for emotion analysis. For eye gaze analysis, SVM was selected, and Openpose was used for body pose analysis, including skeleton detection and calculation of movement or joint angles to detect specific movements.

## 4 System Scenario

### 4.1 System Scenario and Interface

The interactive robot process in this study consists of the Asus Zenbo robot, NVidia Jetson Nano, and a video camera. Zenbo is responsible for the interaction process with children, which serves as the data for subsequent emotion and eye gaze analyses. Meanwhile, the video camera connects to the Jetson Nano to detect and recognize body pose movements from video images. The overall system setup is shown in Fig. 6. During the process, the medical team will observe from the side and fill out the observation scale designed for this system and the medical-specific observation scale required during the diagnosis process.



**Fig. 6.** System and application setup scenario

## 5 Conclusions

This study aims to develop a robot-interactive process and intelligent image analysis system for assisting in the diagnosis of autism in children. Through discussions with professional doctors and references to relevant assessment scales and diagnostic manuals for autism, a diagnostic process has been established on the robot, including calling names, telling stories, singing and dancing, and imitating games. The participant's image information and body pose movements are recorded during the process, and medical staff is asked to fill in the information during the diagnosis process. Finally, this information is presented to the doctor as a diagnostic reference in the form of data visualization. Experimental comparison of the models used in the system. In terms of emotion analysis models, the five models used in this experiment seem to have little difference in the average values of Precision, Recall, and F1-score. Among them, Xception performed relatively well in terms of Precision and F1-score, so Xception was ultimately adopted as the model used for emotion recognition in the system. The eye gaze analysis model uses SVM with average analysis results obtained in different eye gaze directions. The proposed assistive diagnostic robot interaction process and intelligent image analysis system in this study serve as a preliminary combination of medical and technology. In addition to improving the models used in the process, the collected data can also be used to train a classification model for better predicting the probability of autism symptoms and presenting the diagnosis to the doctor in a visualized data format.

**Acknowledgement.** This research was supported by the National Science and Technology Council, Taiwan, under Grant 109-2410-H-197-002-MY3 and 112-2410-H-197-002-MY2.

## References

1. American Psychiatric Association and American Psychiatric Association (eds.). Diagnostic and statistical manual of mental disorders: DSM-5, 5th ed. American Psychiatric Association, Washington, D.C (2013)
2. Lord, C., et al.: Autism spectrum disorder. *Nat. Rev. Dis. Primer* **6**(1), 5 (2020). <https://doi.org/10.1038/s41572-019-0138-4>
3. Al-Dewik, N., et al.: Overview and introduction to autism spectrum disorder (ASD). In: Essa, M.M., Qoronfleh, M.W. (eds.) *Personalized Food Intervention and Therapy for Autism Spectrum Disorder Management*. AN, vol. 24, pp. 3–42. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-30402-7\\_1](https://doi.org/10.1007/978-3-030-30402-7_1)
4. Emanuel, R., Weir, S.: Catalysing communication in an autistic child in a LOGO-like learning environment. In: *Proceedings of the 2nd Summer Conference on Artificial Intelligence and Simulation of Behaviour*, pp. 118–129 (1976)
5. Shamsuddin, S., Yussof, H., Ismail, L.I., Mohamed, S., Hanapiah, F.A., Zahari, N.I.: Initial response in HRI- a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot NAO. *Procedia Eng.* **41**, 1448–1455 (2012). <https://doi.org/10.1016/j.proeng.2012.07.334>
6. Silvera-Tawil, D., Bradford, D., Roberts-Yates, C.: Talk to Me: The role of human-robot interaction in improving verbal communication skills in students with autism or intellectual disability. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Nanjing, Aug., pp. 1–6 (2018). <https://doi.org/10.1109/ROMAN.2018.8525698>

7. Petric, F., et al.: Four tasks of a robot-assisted autism spectrum disorder diagnostic protocol: first clinical tests. In: IEEE Global Humanitarian Technology Conference (GHTC 2014), pp. 510–517 (2014)
8. Petric, F., Kovačić, Z.: Hierarchical POMDP framework for a robot-assisted ASD diagnostic protocol. In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 286–293 (2019)
9. Moghadas, M., Moradi, H.: Analyzing human-robot interaction using machine vision for autism screening. In: 2018 6th RSI International Conference on Robotics and Mechatronics (ICRoM), Tehran, Iran, Oct., pp. 572–576 (2018). <https://doi.org/10.1109/ICRoM.2018.8657569>
10. Javed, H., Park, C.H.: Behavior-based risk detection of autism spectrum disorder through child-robot interaction. In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge United Kingdom, Mar., pp. 275–277 (2020). <https://doi.org/10.1145/3371382.3378382>
11. Pruette, J.R.: Autism diagnostic observation schedule-2 (ADOS-2). Google Sch., pp. 1–3 (2013)
12. McCrimmon, A., Rostad, K.: Test review: autism diagnostic observation schedule, second edition (ADOS-2) manual (Part II): toddler module. *J. Psychoeduc. Assess.*, **32**(1), 88–92 (2014). <https://doi.org/10.1177/0734282913490916>
13. Zahara, L., Musa, P., Prasetyo Wibowo, E., Karim, I., Bahri Musa, S.: The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based raspberry Pi. In: 2020 Fifth International Conference on Informatics and Computing (ICIC), pp. 1–9 (2020). <https://doi.org/10.1109/ICIC50835.2020.9288560>
14. FER-2013. <https://www.kaggle.com/datasets/msambare/fer2013>. Accessed 29 June 2022
15. Florea, L., Florea, C., Vrânceanu, R., Vertan, C.: Can Your Eyes Tell Me How You Think? A Gaze Directed Estimation of the Mental Activity (2013)
16. Vrânceanu, R., Florea, C., Florea, L., Vertan, C.: NLP EAC recognition by component separation in the eye region. In: International Conference on Computer Analysis of Images and Patterns, pp. 225–232 (2013)
17. Pedregosa, F., et al.: Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**(85), 2825–2830 (2011)
18. Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291–7299 (2017)
19. Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y.: OpenPose: realtime multi-person 2D pose estimation using part affinity fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(1), 172–186 (2019)
20. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). <https://doi.org/10.1109/CVPR.2009.5206848>