Applied Mathematics and Computational Intelligence
Volume 13, No. 2, 2024 [1-10]

A Deep Learning Approach for Face Detection and Recognition to Initiate Human-Robot Conversation

Abdul Halim Ismail[1*], Mohammed Khaled Ahmed Al Ghaili[1], Mohamad Amir Hamzah Md Yusof[1], Saeed Akash Mastoi[1], Muhammad Hisyam Rosle[1], Bukhari Ilias[1] and Muhamad Safwan Muhamad Azmi[2]

[1]Dept. of Mechatronic Engineering, Faculty of Electrical Engineering & Technology, University Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, MALAYSIA
[2]Faculty of Mechanical Engineering & Technology, University Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, MALAYSIA.

[*]Corresponding author : ihalim@unimap.edu.my

Received: 16 January 2024
Revised: 31 January 2024
Accepted: 12 March 2024

ABSTRACT

Artificial Intelligence (AI) is currently booming at almost all field. The inauguration of OpenAI ChatGPT using Natural Language Processing (NLP) has played a vital role in exposing AI to the public. It is estimated about 1.8 billion users visit ChatGPT site in a month, with further planning of apps creation in iTunes Apple App Store and Android Google Playstore. Therefore, it is interesting and natural to implement such technology in robotic field. This paper presents the attempt to employ AI into the mobile robot system towards the main goals of conversational intelligence between human and robot. First, the robot head is designed and assembled, then a screen that functioned as the robot face is attached. Afterwards the detection and recognition system were developed giving the ability to the robot to recognize registered persons and the robot eye is able to track where the person is, in the camera Field-of-View (FOV). In addition, all these systems are developed on in-situ device i.e. NVIDIA® Jetson™ Nano. It is targeted that the proposed system is able to initiate a natural conversation between a robot and a human user.

Keywords: Mobile robot, conversational intelligence, deep learning.

1 INTRODUCTION

Service robots are a subset of mobile robots which are designed to perform tasks that assist or augment human activities. They can be used in various settings, including homes, hospitals, and offices. Service robots can be programmed to perform multiple tasks, including cleaning, cooking, and assisting the elderly or disabled. Some examples of service robots include vacuum cleaners, lawnmowers, personal assistants [1], and recently famous waiter robot used in the restaurant [2]. A service robot that shares human characteristics i.e. face such as facial expression will present a more welcoming feeling during an interaction with a human user. Currently, advanced humanoid robots use at least 52 motors for movements and expressions, followed by a synthetic material to create robot skin [3]. Some of the recent robot face is developed on an interactive screen which cost much

less compared to the animatronics faces. Figure 1 shows the difference between animatronic and screen-based robot face.

Conversational intelligence (CI) is a feature of a service robot which is aimed to converse naturally with a human user. A robot having CI, for example Sophia is a real-life example of a social robot employing CI [4]. One interview with Hanson Robotics's Sophia prompting about future in social robotics yield in the answer of "I think developments in natural language processing will have a big impact in customer service, too". It is yet interesting and human are cautiously and eagerly observing the robot capabilities so that the human doom via robotic catastrophic can be avoided.

Figure 1: (Left) Animatronic robot face, Sophia [5] and (right) Screen-based robot face [6].

At present CI plays an interesting factor to catch human attention in human-robot interaction. CI technology consists of several components in order to be realized. Figure 2 shows the system architecture for a typical conversational intelligence system [7]. The components of CI begin with automatic speech recognition (ASR) to convert speech input from the user into text, Natural Language Understanding (NLU) to analyze the text and determine its meaning, a dialogue management system that interacts with knowledge sources to determine the system's action, and Natural Language Generation (NLG) and text-to-speech (TTS) components that convert text into speech for the user. It is worth noting that the architecture for a text-based conversational AI system would differ from this because it would not include ASR and TTS [7].

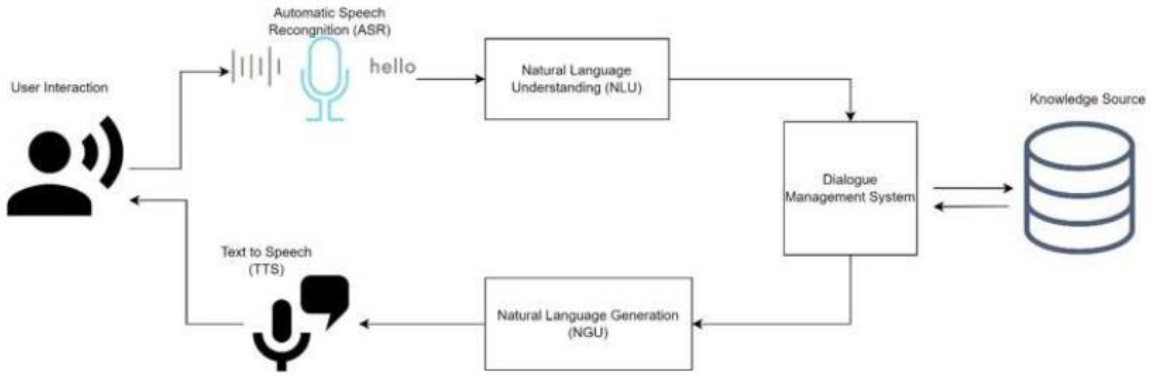


Figure 2: An example of CI architecture in [7].

An intelligence social robot interaction with humans employing face recognition by tracking the user for initiating the conversation is typically based on vision and sound sensors [8, 9]. There are

numerous commercial applications using such technologies for instance in studying the social gaps using robotics [10] and scientific interest for learning [11]. In general, the robot is expected to know whom its converse with before initiating the conversation. This would ensure that the conversation is meaningful and able to provides exhilarating memories to the user. Many literatures use advanced machine vision methods to detect facial expressions, eye gazing, facial location as well as head post, often by deep learning-based Convolutional Neural Network (CNN) [12]. For facial features or face recognition, the input data of and many video surveillance into the machine learning utilizing various computer vision algorithms such as expression detection, face detection and so on for identification of person [13]. This paper explores a feasible approach for the robot to identify who is the user by means of face detection and face recognition. The artificial intelligence tools embedded in the Nvidia Jetson nano is used. The screen-based robot face is also encapsulated in an in-house robot head design featuring aesthetic design elements. The robot head is firstly simulated and then were later printed in a metal sheet before putting the screen for the robot face.

2 ROBOT HEAD DESIGN

A social robot would be more aesthetic and appealing with robot head. In the case of our robot, we have design and develop an aesthetic robot head using Autodesk Inventor software. Then, the design was printed to a 3mm metal sheet and cut via laser cutting. It was later folded, and spot welding was used giving nice finishing to the head. Figure 3(a-d) shows the design, the finishing as well as assembled robot head with the 7-inch screen which later will be the robot face. Aside from the screen, the robot head is also prepared spaces for additional components such as for overhead LIDAR, speakers, microphone and wiring area. Further works will emphasize these matters.

(a) (b) (c) (d)

Figure 3: Design and assembly of the robot head, (a) Robot body without the head, (b) The designed robot head, (c) Actual robot head and (d) Fully assemble robot head and screen.

3 ROBOT FACE DESIGN

The eyeball of the robot face is used as graphical medium to track the user location in the camera field-of-view (FOV). The eyeball movement is controlled by computing the coordinate of the user's

face resulting from face recognition algorithm. The coordinate is updated with respect to the location. The center of the user's face s given as

$$x_c = \frac{2x + w}{2} \tag{1}$$

$$y_c = \frac{2y + h}{2} \tag{2}$$

where x_c and y_c are the center of user's face, x and y are the current coordinates, and w and h are the width and height of a bounding box of the detected face respectively. The distance and angle between the eyeball coordinates and the user's face center are given as

$$\begin{aligned} x_{dist} &= x_c - x_{eye} \\ y_{dist} &= y_c - y_{eye} \\ d &= \min\left(\sqrt{x_{dist}^{\ 2} - y_{dist}^{\ 2}}\right) \\ \theta &= \tan^{-1}\left(\frac{y_{dist}}{x_{dist}}\right) \end{aligned} \tag{3}$$

The pupil location of the robot eyeball is then given as

$$\begin{aligned} x_{pupil} &= x_{eye} + \left(d\cos\theta\right) \\ y_{pupil} &= y_{eye} + \left(d\cos\theta\right) \end{aligned} \tag{4}$$

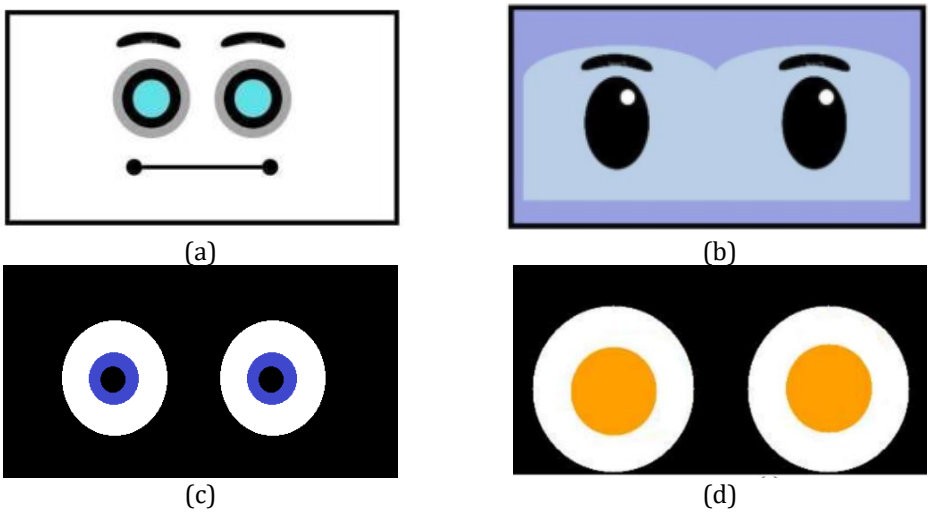


Figure 4: The design of robot face on screen, (a) Initial design in GIMP, (b) anime-like design in GIMP, (c) Initial design in OpenCV and (d) revised design in OpenCV.

4 ROBOT FACE DETECTION

4.1 Face Detection Method

The Haar cascade based on Viola-James algorithm was primarily used to detect the faces from the images that were captured using the camera mounted on the robot head. The Haar Cascade algorithm is utilized for object detection in images by employing features extraction method. A cascade function is trained using a large dataset comprising both positive and negative images, in our case the images of human face. The training of Haar cascade is based positive images which contained images of face and negative images with zero human face. The Haar kernel is then convoluted with the training images to find a face in an image comparing them to the training images. The advantage of this algorithm lies in its ability to achieve real-time performance without the need for computationally intensive calculations. Haar cascade is also chosen because of the simplicity and efficacy of Haar-like features where the detection speed has neared the goal of practical application [18].

In this works, we begin the face detection by the insertion of a Haar cascade XML file specifically designed for human faces. The images are later obtained from the robot's camera in frames. The frames' complexities are subsequently reduced using image processing techniques. This is achieved by transforming the color image into grayscale, and resizing it, preparing for next phase. The Haar cascade technique is then used to extract characteristics from the processed frames and determine the existence of any faces. When a user face is successfully identified, a bounding box is formed around the detected face. The bounding box is defined by coordinates defined before as xc, yc, w and h. Figure 5 shows the detected human face with the bounding box in order.

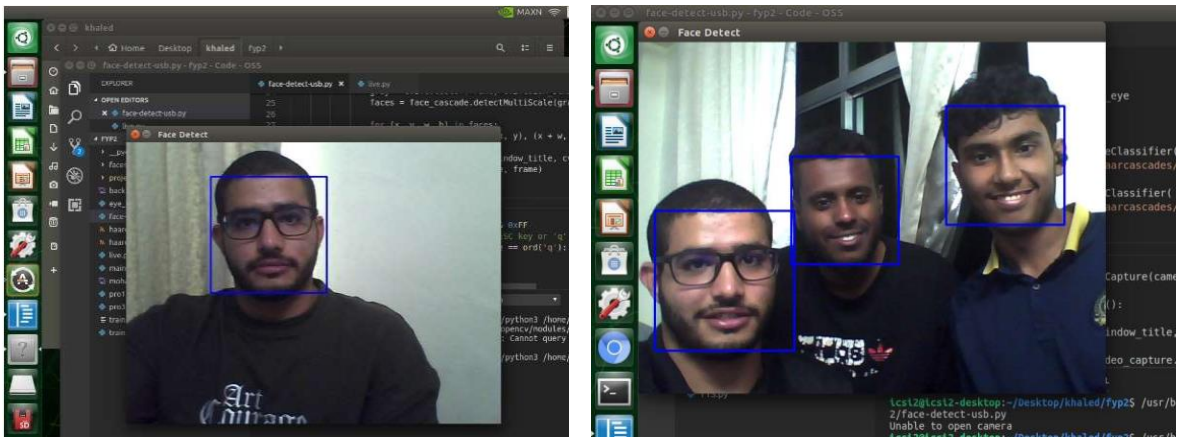


Figure 5: The detected human face using Haar Cascade algorithm in Nvidia Jetson Nano device with the bouding box (left) single user (right) multiple users

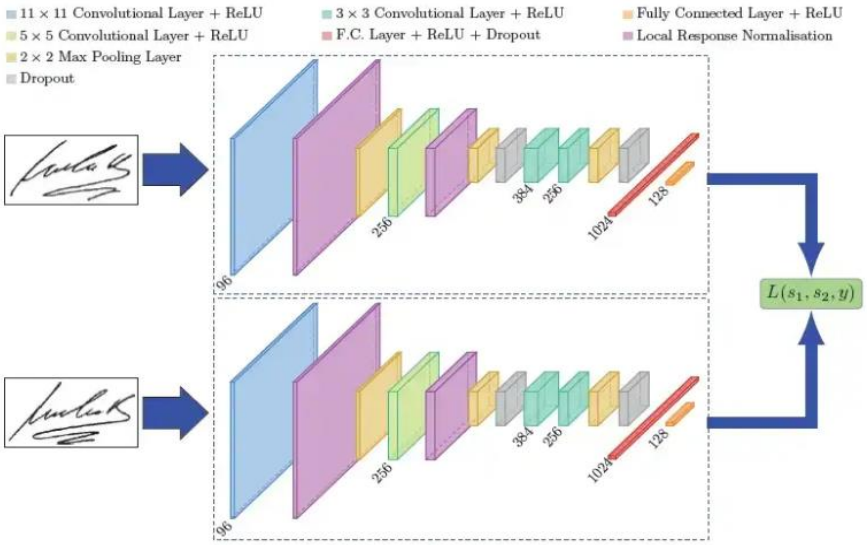


Figure 6: An example of CNN architechture of SigNet deep learning algorithm [19]

4.2 Face Recognition Method

When a user face has been detected, the next step is to apply face recognition so that the robot acknowledges the person in front of it in order to initiate conversation. In this works, the Dlib's ResNet-based face recognition model algorithm is employed to recognize the user face who are detected in the previous face detection using Haar Cascade method. Similar to Haar cascade algorithm, Dlib's works by training both positive (known user) and negative (no known user) image in a deep neural network architecture typically using Convolutional Neural Network (CNN) as shown in Figure 6 [19]. Once the network is trained, it can then be used to generate embeddings for new images. These embeddings can then be used to train a classifier to recognize faces. The classifier can be any machine learning algorithm, such as K-Nearest Neighbors (KNN), support vector machine (SVM), Random Forest, etc. Hence, the face is recognized when the data with the closest embedding is found.

Following that, the model is tested by presenting it with a facial image of the same individual. This image is then compared to the known faces stored during the training phase by the model. If the face is correctly detected, a rectangular outline is created around it, and the name of the individual is written within the rectangle zone. This can be done by using the same coordinates that were used to draw the rectangle outline. Figure 7 shows the human recognized with the name of the user is embedded on top of the bounding box.

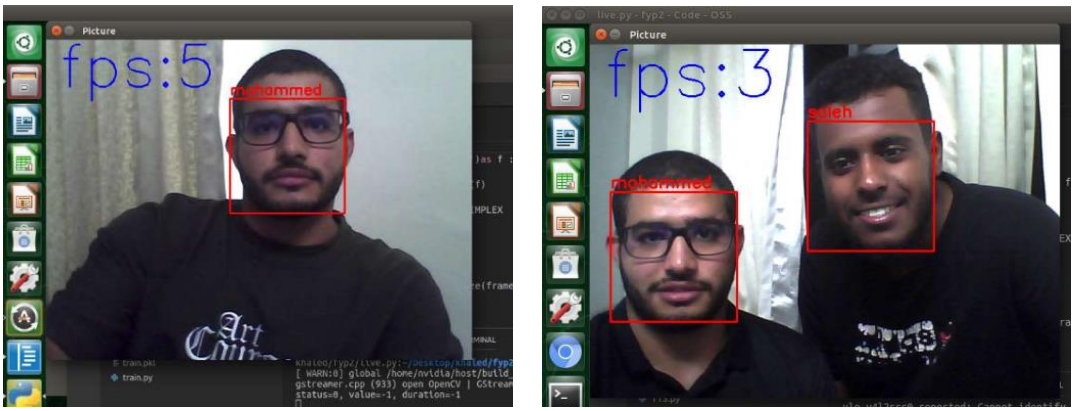


Figure 7: The recognized human face in front of the robot in Nvidia Jetson Nano, (left) Single user and (right) multiple users.

5 RESULTS AND DISCUSSION

5.1 Face Detection and Recognition

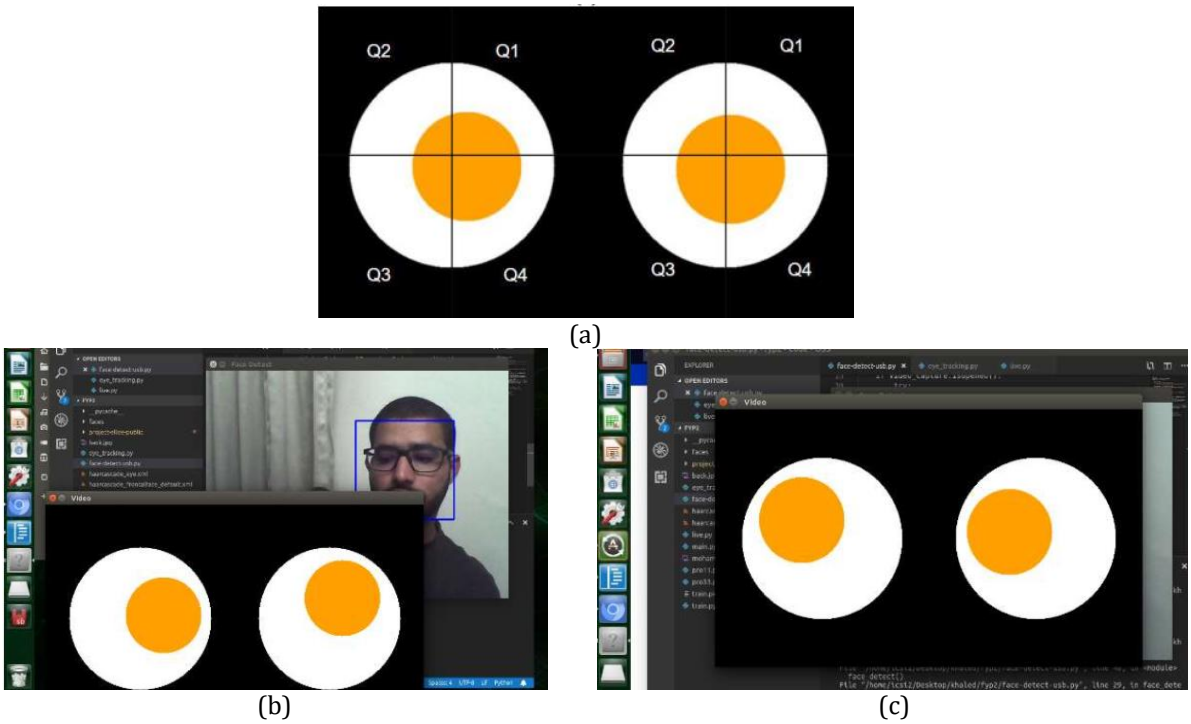
The main goal presented in this paper is to analyze the robot ability to recognize the human and the capability of the robot eyeball on a robot face to track the user in front of it before initiating the conversation. Therefore, the analysis conducted in this paper is to observe the maximum distance between the user and the robot before recognition is lost. A distance-based observation is made, where the human is standing on front of the robot of specific distances. The measurement of the area of bounding box is computed to ensure that the face is detected in the robot FOV. Table 1 tabulated the results of the face detection and face recognition. It can be observed that the face can be detected regardless of any distances, but the face is unable to recognize at 2m distance and forward. This is mainly because the face is quite away from the robot making the recognition fail. At 0.5m it is found that the box is too big, making some face features missing. Therefore, the maximum distance to initiate the conversation between the robot and human is 1.5m.

Table 1: Result of face detection and recognition

Distance (m)	Area of bounding box (pixels)	
	Face Detection	Face Recognition
0.5	87024	-
1.0	18088	30100
1.5	12916	20878
2.0	7744	-
3.0	2862	-

5.2 Human Face Tracking

The robot face as depicted in Figure 4(d) was chosen to be embedded with the face detection and recognition method. In order to observe the correct user face tracking from the robot's eyeball point of view, an analysis is made based on quadrant was made. Figure 8 shows the quadrant setting and the resulting user face tracking.



(a)

(b) (c)

Figure 8: The face tracking analysis, (a) the setting of robot eyeball quadrant, (b) the eyeball succesfuly track user at Q1, and (c) the eyeball succesfully track user at Q2

6 CONCLUSION

The works on developing the functional robot head and the robot face with face detection and recognition functionality on a Nvidis Jetson Nano device provides important insights and contributions to the field of robotics. Such technology is expected to assist the whole objective in developing the natural conversational intelligence for a service or social robot. The study represents a step towards more complex and socially adept mobile robots, which will eventually bridge the communication and interaction gap between humans and robotics.

The proposed system incorporates various technologies to achieve its functionality. The deep learning approach particularly for face detection utilizing Haar Cascade algorithm and face recognition using Dlib Dlib's ResNet-module are able to detect user faces in real-time, resulting in high accurate and quick detection. Hence it should be able to be used in actual application. Simple measurements were made and the maximum distance for face detection and recognition is at 1.5m from the robot distance.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support from the Centre of Excellence for Intelligent Robotics & Autonomous Systems (CIRAS), Universiti Malaysia Perlis.

REFERENCES

[1] M. Veloso, J. Biswas, B. Coltin, and S. Rosenthal, "CoBots: Robust Symbiotic Autonomous Mobile Service Robots", Proceedings of the 24th International Conference on Artificial Intelligence, AAAI Press. pp. 4423, 2015.

[2] S. Sotnik, and V. Lyashenko, "Prospects for Introduction of Robotics in Service", *International Journal of Academic Engineering Research (IJAER)*, vol. 6, no. 5, pp. 4-9, 2022.

[3] C. Wagner, "Silver Robots' and 'Robotic Nurses? Japanese Robot Culture and Elderly Care Demographic change in Japan and the EU", Comparative perspectives, pp. 131-54, 2010.

[4] J. S. Retto, "First citizen robot of the world", Accessed on ResearchGate URL: https://www. researchgate.net, 2017.

[5] A. Aly, S. Griffiths, and F. Stramandinoli, "Metrics and Benchmarks in Human-Robot Interaction: Recent Advances in Cognitive Robotics", *Cognitive Systems Research*, pp. 13-23, 2017.

[6] A. Johnson, T. Roush, M. Fulton, and A. Reese, "Implementing Physical Capabilities for an Existing Chatbot by Using a Repurposed Animatronic to Synchronize Motor Positioning with Speech", *International Journal of Advanced Studies in Computers, Science and Engineering*, vol. 6, no. 1, pp. 20, 2017.

[7] A. B. Saka, L. O. Oyedele, L. A. Akanbi, S. A. Ganiyu, D. W. Chan, and S. A. Bello, "Conversational Artificial Intelligence in the AEC Industry: A Review of Present Status, Challenges and Opportunities", *Advanced Engineering Informatics*, vol. 1, no. 55, pp. 101869, 2023.

[8] C. Sirithunge, A. B. Jayasekara, and D. P. Chandima, "An Evaluation of Human Conversational Preferences in Social Human-Robot Interaction", *Applied Bionics and Biomechanics*, pp. 1-3, 2021.

[9] C. Lee, Y. S. Cha, and T. Y. Kuc, "Implementation of dialogue system for intelligent service robots", *IEEE International Conference on Control, Automation and Systems*, pp. 2038-2042, 2008.

[10] K. Kühne, M. A. Jeglinski-Mende, M. H. Fischer, and Y. Zhou, "Social Robot–Jack of All Trades?", *Paladyn Journal of Behavioral Robotics*, vol. 13, no. 1, pp. 1-22, 2022.

[11] J. E. Michaelis, and B. Mutlu, "Supporting Interest in Science Learning with a Social Robot", *18th ACM International Conference on Interaction Design and Children,* pp. 71-82, 2019.

[12] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. P. Morency, "Openface 2.0: Facial Behavior Analysis Toolkit", *IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 59-66, 2018.

[13] R. C. Damale, and B. V. Pathak, "Face Recognition Based Attendance System Using Machine Learning Algorithms", *IEEE International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 414-419, 2018.

[14] A. Poberznik, and M. Sari, "13 Older Adults' Experiences with Pepper Humanoid Robot", Tutkimusfoorumi, 2019.

[15] P. Barros, W. Stefan, and S. Alessandra, "Towards Learning How to Properly Play UNO with the iCub Robot", arXiv preprint arXiv:1908.00744, 2019.

[16] C. Kertész, and T. Markku, "Exploratory Analysis of Sony AIBO users", *AI & Society*, vol. 34, pp. 625-638, 2019.

[17] C. M. Jones, and A. Deeming, "Speech Interaction with an Emotional Robotic Dog", 9th Annual Conference of the International Speech Communication Association, 2008.

[18] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, "Human Face Detection Algorithm Via Haar Cascade Classifier Combined with Three Additional Classifiers", *IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, pp. 483-487, 2017.

[19] S. Dey, A. Dutta, J. I. Toledo, S. K. Ghosh, J. Lladós, and U. Pal, "Signet: Convolutional Siamese Network for Writer Independent Offline Signature Verification", arXiv preprint arXiv:1707.02131, 2017.

[20] A. H. Ismail, Y. Mizushiri, R. Tasaki, H. Kitagawa, T. Miyoshi, and K. Terashima, "A Novel Automated Construction Method of Signal Fingerprint Database for Mobile Robot Wireless Positioning System", *International Journal of Automation Technology*, vol. 11, no. 3, pp. 459-71, 2017.