

# Smile Detection Tool using OpenCV-Python to Measure Response in Human-Robot Interaction with Animal Robot PARO

Winal Zikril Zulkifli<sup>1</sup>, Syamimi Shamsuddin<sup>2</sup>, Fairul Azni Jafar<sup>3</sup>  
Fakulti Kejuruteraan Pembuatan  
Universiti Teknikal Malaysia Melaka  
Hang Tuah Jaya, Melaka, Malaysia

Rabiah Ahmad<sup>4</sup>  
Fakulti Teknologi Maklumat dan Komunikasi  
Universiti Teknikal Malaysia Melaka  
Hang Tuah Jaya, Melaka, Malaysia

Azizah Abdul Manaf<sup>5</sup>, Alaa Abdulsalam Alarood<sup>6</sup>  
Department of Computer Science,  
Faculty of Computing and Information technology  
University of Jeddah, Jeddah

Lim Thiam Hwee<sup>7</sup>  
Psychosocial Department  
SOCSCO Tun Razak Rehabilitation Centre (TRRC)  
Hang Tuah Jaya, Melaka, Malaysia

**Abstract**—Human-robot interaction (HRI) is a field of study that defines the relationship between humans and robot. In robot-assisted mental healthcare, there is still a lack in the methodology especially in evaluating the outcome. In this study, PARO; a robot in the shape of a cute baby seal is introduced as an adjunct therapy tool for six rehabilitation patients with post-stroke depression. Currently, the therapy outcome is measured using psychological tools. When a robot is introduced, a new measurement tool is needed to analyse the patients' response. Thus, this study constructs a tool using OpenCV-Python to detect the number of smiles when each patient interact with PARO. Smile is an indicator of positive emotion and that PARO helps to uplift patient's mood. The results were then compared with psychological evaluations. Both tools show congruent results. The number of smiles increased when patients were holding PARO and PARO helped all patients to manage their psychological distress. This indicates that smile detection is an effective supporting tool to indicate respond in human-robot interaction.

**Keywords**—Human-robot interaction; OpenCV; PARO

## I. INTRODUCTION

Application of robots in healthcare setting are on the rise. These include robots used in surgery, to carry medicines, exoskeletons and robots for mental rehabilitation. Human-robot interaction (HRI) is a field that looks into the technicalities and relationship between humans and robots. HRI covers how humans interact with robots, and how the robotic systems are implemented in accomplishing tasks in human environments. HRI aims for effective and safe interaction for both entities. As HRI studies cover humanities, engineering, computer science, medical and social science; this makes HRI inherently a multi-disciplinary field of research.

This project is funded under the Fundamental Research Grants Scheme (FRGS) [FRGS/1/2016/SKK06/FKP-AMC/ F00321].

To date, robots have been used in therapy to treat patients with mental disabilities such as autism, Alzheimer and dementia [1-3]. With robotics involvement, it is expected that the condition of people who suffer from mental disabilities can be improved [3]. Psychotherapy alone may yield positive results, but combination treatments have the best outcome. This is why robotic therapy is suitable as it can be integrated as a tool in the current therapy program

To evaluate the outcome human-robot interaction, a few methods are usually used including surveys, interviews and bio sensing tools. For robotic intervention involving patients with depression, a specific tool is needed to assess the respond through the recorded videos when patients interact with the robot. This is to complement the current psychological tools used by therapists to assess patient conditions. Interaction can be recorded and the videos can be analysed during post-processing stage. This method does not involve asking questions to patients or putting devices on the patient's body.



Fig. 1. PARO the Seal Robot.

Machine learning is a method that allows systems to learn and improve from data and make decisions. In this study, a smile detection tool using OpenCV-Python was developed to detect the number of smiles when each patient interacts with robot. Through machine learning, the number of smiles are detected and serves as an indicator to recognize human's mood while interacting with the robot. This is based on the understanding that the generation of a real smile is produced when a person is feeling positive. Smiles are detected to indicate the mood of a person when they interact with the robot. More smiles means the person's mood is improving.

In this HRI study, an animal-robot PARO (Fig. 1) was used during therapy for patients with post-stroke depression at a multi-disciplinary rehabilitation centre. Most of previous researches have focused the use of PARO robot to treat patients with dementia [4]. There was also a study that investigated which function of PARO that can be associated to reduce psychophysiological stress [5]. PARO was chosen because of its cute and appealing look. It is commercially available, its weight is suitable to patients to hold and hug it and it is the only robot classified as a class 2 medical device by the U.S. Food and Drug Administration (FDA).

The aim of this study is to construct a tool to detect the number of smiles from recorded videos of interaction involving patients with depression and animal-robot PARO. The results of the number of smiles were compared to psychological assessments by a psychologist to evaluate its suitability in the current HRI study.

## II. METHODOLOGY

The research methodology is shown in Fig. 2. The process starts with developing the research protocol that involves human and robot interaction. The model was based on literature review, subjective evaluation with public respondents in Malaysia and advice from experts in rehabilitation psychology and HRI. Approval from research ethic committee in Universiti Teknikal Malaysia Melaka (UTeM) was filed based on the experimental model. Permission was granted on 19<sup>th</sup> October 2016. The approval is important because this experiment involved humans as subjects and their rights, as well as the rights of the researchers need to be protected.

The research continued with OpenCV-Python program development for smile detection. This program was developed as a HRI evaluation tool to analyse the outcome of the experiment. Next step was the full-scale HRI study with six patients in the age between 35 to 52 years old. Two were males. All patients were diagnosed with post-stroke depression. The inclusion criteria were: diagnosed with depression after stroke, currently an in-patient at the SOCSO Tun Razak Rehabilitation Centre (TRRC) Melaka, able to converse in Malay language or English, able to stroke an animal and can hear clearly. They were currently undergoing rehabilitation therapy program at TRRC. Patients with depression at centres like TRRC undergo rehabilitation treatment before they are declared stable.

The set-up of the session involving patient, PARO and psychologist is shown in Fig. 3. The experiments were conducted in three separate sessions. First, the therapy room in

TRRC were set-up with HD video camera, voice recorder and projector. Next, the patient was invited to enter the room. Then, the patient was given brief introduction about the experiment, his/her consent was taken, and a short video of PARO was shown.

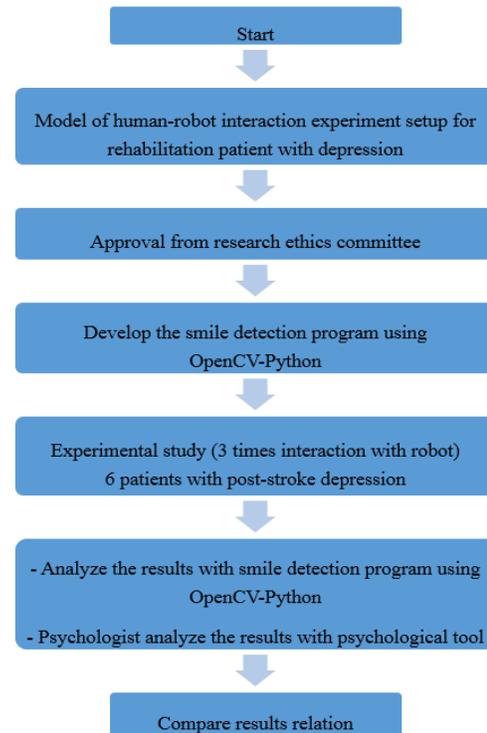


Fig. 2. Flowchart of the research methods involved in this study.

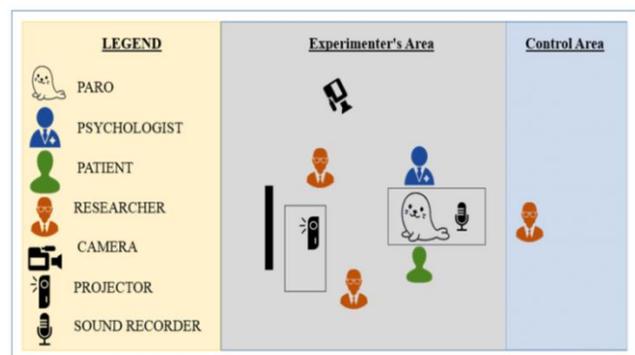


Fig. 3. Experimental layout.

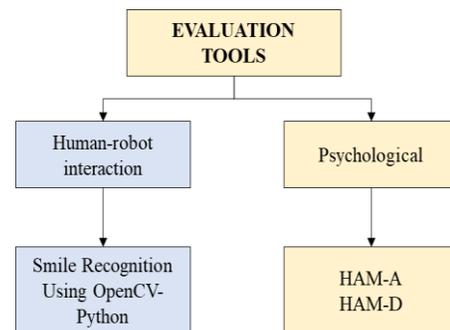


Fig. 4. Evaluation tools used in the study.

The following step was the psychological pre-evaluation on the patients by a rehabilitation psychologist from TRRC using the Hamilton Depression Rating Scale (HAM-D) and Hamilton Anxiety Rating Scale (HAM-A). HAM-D is a validated screening instrument for post-stroke depression [6] to provide indication of depression, guilt, suicide, insomnia (sleep problem) and as guide to evaluate recovery. HAM-A measures the severity of anxiety symptoms and is widely used in both clinical and research settings. The reason is to identify the patients' condition before robotic intervention and identify the changes in patients in respond to the robot.

Then, the 15 minutes session took place to obtain therapy recording without PARO. During this period, the psychologist began to introduce general topics to talk with the patient such as their favourite food, their activities during weekends and their family members. Indirectly, the psychologist was using cognitive behaviour techniques and goal directed behaviour during the therapy. After 15 minutes, interaction with PARO began. PARO was brought into the room by the experimenter and was put in front of the patient. With PARO in the room, the consultation therapy continued with PARO as a companion to the patient. The psychologist continued to have conversations with the patient. The 15 minutes therapy without PARO and 15 minutes with PARO cycle was repeated for all three sessions. At the end of each session, the patient was informed that PARO will be taken away. Total duration of recorded experiment was limited to 30 minutes because post-stroke patients cannot withstand long therapy sessions. They also get tired easily.

The interaction took place at a therapy room in TRRC for 3 sessions during one month. Each session lasted for 30 minutes and it included therapy time with PARO and without PARO. For the second and third session, the psychologist also carried out the HAM-D and HAM-A evaluations. The same protocol was applied for all the six patients recruited for the study.

#### A. Evaluation Tools: HRI and Psychological

To evaluate the outcome of experiment, the tools are used that are separated into two categories: human-robot interaction (HRI) and psychological evaluation as shown in Fig. 4. HRI evaluation consists of the smile recognition tool developed using OpenCV-Python. A smile is a positive emotion. To identify the changes in depression patients, smile has been chosen as an indication of the patient's positive mood. To identify the smile occurrences, the OpenCV-Python program analysed the recorded video from the patient-robot interaction session.

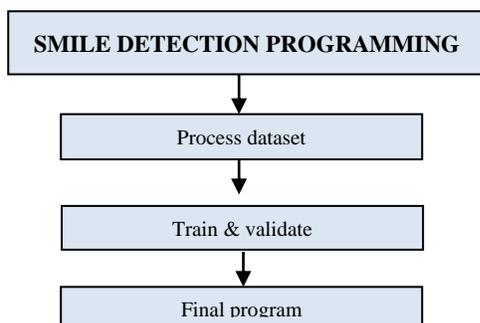


Fig. 5. Smile detection development program flow.

For the psychological tools, HAM-D and HAM-A were chosen because of their reliability and recommendations from the psychologist. HAM-D and HAM-A were used together because the observations on patients with depression always correlate with assessments of depression and anxiety.

#### B. Smile Detection Programming using OpenCV-Python

The development phase of the OpenCV-Python is shown in Fig. 5. The objective of the program is to identify and count the smile on patients during the experiment. First, a collective data on human smiles must be included for smile identification. This action followed by the second step which is identifying appropriate dataset processing library. OpenCV is an open source computer vision library for commercial and research use. OpenCV is one of the most widely used libraries in image processing. The OpenCV was chosen for its extensive library, it is simple to use and has extensive user network.

Once the dataset library had been identified, the development continued with detection and learning. For smile detection algorithm, Adaptive Boosting (AdaBoost) was used. Using AdaBoost improves the accuracy of the learning algorithm where the output of multiple "weak classifiers" are combined into a weighted sum that represents the final output of the boosted classifier.

Smile detection was targeted at the mouth region. In the smile detection mechanism, the videos were processed frame by frame and converted into a grey scale. Smile detection determines 'where' in a face image the 'smile' is located, and this is done by scanning the different face image scales and extracting the exact patterns to detect a smile. The classifier uses a single feature to define images as smiles or non-smiles which are stored in cascade data of the numbers of 'smiles'. The program is trained using the pre-determined dataset in OpenCV for program environment adaptation. The program analysed the recorded videos from the experiment.

To prepare a dataset, raw images of smiling faces were needed. Previously, Chang et al. used ten individual subjects and captured four images from each of them [7]. This resulted with 40 sample images for the database.



Fig. 6. Sample dataset of 300 smiling faces of Malaysians.

However, in this study, the developed program in OpenCV-Python is able to capture ten images within 100 milliseconds. These images were acquired by running a video through an OpenCV-Python program to capture still images. Using video recordings of thirty different individuals, 10 smiling images were extracted from each individual. This makes a total of 300 smiling faces (Fig.6) for the database. The steps to run this program are divided into three main program sections which are: detector, dataset creator and trainer. For performance validation, the dataset was divided into training and validation. The verification of this data used the 5-fold cross validation [8]. Then, the output images were stored using unique identification codes. The images are labelled with target vector '1' for smiling and '0' for not smiling. Images that are labelled with not smiling are used as negative image dataset.

The HRI smile detection tool was made with OpenCV-Python program to introduce a new dataset with acceptable accuracy for smile identification. With the new dataset, the accuracy of the HRI smile detection program was improved. The new dataset also introduced sample images from Malaysians.

### III. RESULTS AND DISCUSSION

In this study, the HRI smile detection program was used to analyse the video recordings. The performance of the program was tested before it was applied in the experiment study.

#### A. Performance of the Smile Detection Program

To evaluate the effectiveness of this program, an accuracy test was carried out. The test used three types of dataset for tracking. Each dataset consists of pictures of human faces. 'Smile' pictures were labelled with '1'. The 'not smile' pictures were label with '0'.

To test for accuracy, dataset of images was fed into the smile detection program. The first test used the dataset of 300 images from 10 Malaysian subjects. To test for accuracy, this dataset used 5-fold cross-validation. The technique performs learning and training at the same time.

The second dataset used to test for accuracy was GENKI. The GENKI database, developed by MPLab is an expanding database of image containing faces spanning from a wide range of illumination conditions, geographical locations, personal identity, and ethnicity [9]. The database contains 25,064 images. The third dataset was the Japanese female facial expression (JAFEE) [10]. The database contains 213 images from 60 Japanese subjects. The results on program accuracy are shown in Table 1.

Dataset of the 300 smiling faces of Malaysians scored the highest 90% accuracy. This high score is expected because the same dataset was used to teach/train the program. The results with JAFEE and GENKI dataset are at 76% and 70%. These scores were acceptable to run smile detection on the videos for the main experiment. In image processing, the acceptable image detection accuracy should be more than 70% [11]. Therefore, this smile detection program developed in this study has acceptable accuracy with more than 70% score in the testing result.

TABLE I. ACCURACY RESULTS FOR DATASET TESTING

No.	Dataset	Accuracy
1	300 smiling faces of Malaysians (5-fold cross-validation)	90%
2	JAFEE	76%
3	GENKI	70%

#### B. Results of Smile Detection for Experiment with PARO

The smile detection program was used to analyse the recorded videos from three sessions for each of the six patients. Fig. 7 shows the results of the smile detection program. For each session, the time is separated into 15 minutes spent without the robot and 15 minutes with the robot.

Patient 1, Patient 3 and Patient 4 have shown a continuous increase of the number in smiles starting from the first session until the third session with PARO. Their smile increases with each session in sequel. For Patient 2, Patient 5 and Patient 6 there was also an increase in the number of smiles from the first to the second session. However, the number of smile drops after the second session. When comparing the number of smiles during sub-sessions spent with the robot and without the robot, it is seen that all patients, except for Patient 1, show an increase in the second sub-session with PARO.

Looking at the results of smile detection for the sessions without PARO, it can be seen that the number of smiles also increase except for Patient 5 and Patient 6. For the other 4 patients, they were feeling happier because they are responding well to their rehabilitation program at TRRC. Nevertheless, the presence of PARO managed to increase their happiness and positive mood as the session with PARO resulted with more smiles than the session without PARO. This is a positive outcome. The results indicate that the animal robot PARO gave a positive impact on them emotionally.

On the other hand, Patient 1 in the second session recorded a higher number of smiles in the period without PARO than with PARO. This interesting finding indicate that though PARO is cute and likeable, there is also a small number of people that do not like the robot or find it attractive. Specifically, for patients undergoing rehabilitation therapy, introducing PARO as a tool during therapy may also relate to the patient's acceptance of a robot considering that this is the first study of its kind in Malaysia

#### C. Results of Psychological Tools

Evaluation using the HAM-D and HAM-A were carried out by the psychologist and the results are tabulated in Table II. When using the HAM-D instrument, the score of 0 to 7 is within normal limit, score of 8 to 13 indicates 'mild severity' and score between 14 to 17 falls under 'moderate severity'.

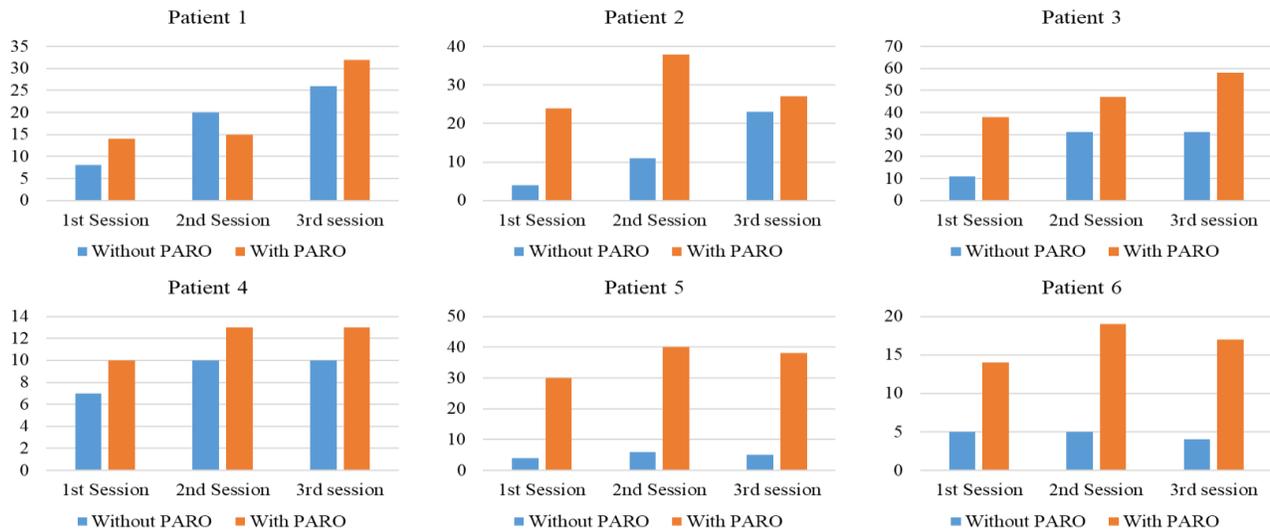


Fig. 7. Results for HRI smile detection program.

Before the experiment, Patient 1 was diagnosed with ‘mild severity’ depression level. After second session, Patient 1’s depression level rise but still in the same level. After the third session, Patient 1 level of depression went down to level ‘within normal limit’. This was better compared to the first evaluation.

Patient 2 and Patient 5 were diagnosed with ‘mild severity’ depression before the experiment started. Their condition showed improvement after second and third session with PARO. Their depression level had improved to ‘within normal limit’.

For Patient 3 and Patient 6, their depression level started with ‘mild severity’ and continued with the same level in the second session. After the third session, Patient 3’s depression level improved to ‘within normal limit’. On the other hand, depression level for Patient 6 also decreased into ‘mild severity’ level. Patient 4 depression level showed no changes, comparing before the experiment started and after the third experiment session, which was in the ‘mild severity’ level. Overall, depression level of all six patients had decreased and their mental health were getting better.

With HAM-A, the score of 0 to 15 is within normal limit, score 16 to 20 is ‘mild severity’ and the score of 20 to 25 falls under ‘moderate severity’. Before the experiment, all patients was screen with HAM-A for assessment on their anxiety level. Patient 1, Patient 4 and Patient 5 level of anxiety were ‘within normal limit’ since before the experiment and until the third session. Even though the Patient 1 showed an increased score anxiety, but the scores were still within the normal level.

On the other hand, Patient 4 and Patient 5 scores had decreased. For Patient 2, her level of anxiety was screened with ‘moderate severity’ before the experiment and after the second session. In the third session, her anxiety level reduced to ‘mild severity’ level. Patient 3 was also screened with ‘moderate severity’ before the experiment. But the anxiety level dropped to ‘mild severity’ level after second and third session. This is in contrast to Patient 6 who was screened with ‘mild severity’ before the experiment and after the second session. After the third session, the level of anxiety fell to level ‘within normal limit’. Overall, all the patients’ score of anxiety had decreased, except for Patient 1. This result shows that most patients are recovering from their anxiety condition.

TABLE II. SCREENING RESULTS OF PSYCHOLOGICAL TOOLS

Screening	HAM-D			HAM-A		
	Pre-experiment	2nd	3rd	Pre-experiment	2nd	3rd
Patient 1	9	13	7	11	14	15
Patient 2	8	5	5	20	21	18
Patient 3	12	9	6	23	18	17
Patient 4	13	13	11	15	11	11
Patient 5	8	6	3	10	8.8	6
Patient 6	15	14	10	18	18	15

D. Relation of HRI Tool with Psychological Tools

To find out whether the results from smile recognition program is congruent with psychological tools assessment, both data were analysed using SPSS software with analysis of variance (ANOVA). The purpose of using ANOVA analysis is to find out the statistical models from the tool’s results and associate their data for estimation through mean square, *F-test* (*F*), *degree of freedom* (*df*) and *significant* (*Sig.*).

Table III shows the results output from ANOVA. Mean square value was used as denominator for *F* ratio. The *F* statistic is defined as the ratio between the two independents. In this study, the independent is between HRI smile recognition tools with HAM-D and HAM-A. The test statistic for *F* value resulted with 9.264, 12.157 and 12.157. Using an  $\alpha$  value of 0.05, we have *Sig.* = 0.001, 0.000 and 0.000. The *df* value of 5 and 12 were applied with these levels.

Since the test statistic is much smaller than the critical value, the null hypothesis is accepted ( $\geq 0.5$ ) of equal population means. This concluded that there is a (statistically) significant among the population means. So, this statistic test proves that the results from HRI smile recognition tool were significant at the predefined *df* levels with the HAM-D and HAM-A tools.

All patients who took part in this study were ordinary (typical) individuals, which means that they were not diagnosed with psychiatric illness. They were admitted to the rehabilitation facility due to their post-stroke condition. Therefore, it is relevant to use a smile detection program with the trained database from ordinary people too. The smiles produced by these patients show congruency with the results from psychological assessment tools. Results were different (in terms of scores) between each patient because, as suggested by the rehabilitation psychologist all patients react differently when they interact with animal robot PARO.

Each patient has their traits. However, in the results with six patients, there are significant between depression and smiles. More smiles mean their mood were improving. PARO give the patients more smiles during the interaction. It is through interacting (i.e., touch, hug, stroke, pet and kiss) and communicating with PARO. This behaviour improves the moods of the patients and as a results, they smiled more. When mood improves, depression is reduced.

Anxiety as a mental condition always have comorbidity results with an individual's depression level. Based on the statistical results level of anxiety have relationship with smiles. In this study, the patients' anxiety level reduced when their number of smile increased. With the presence of PARO during the interaction, it is suggested that the patients were recovering and their positive emotion increased. The patients' mood also improves after each interaction.

Overall, results from HRI smile recognition tool was similar with psychological tools results. Thus, smile recognition can be used along with psychological tools to evaluate the interaction of patients with animal-robot PARO.

TABLE III. ANOVA TEST ON HAM-D AND HAM-A

Tools	df	Mean Square	F	Sig.
HRI (Smile Recognition)	5	577.967	9.264	0.001
	12	62.389		
HAM-D	5	59.869	12.157	0.000
	12	4.924		
HAM-A	5	59.869	12.157	0.000
	12	4.924		

#### IV. CONCLUSION

The study aim was achieved by the development of a smile recognition tool using OpenCV-Python for use in recorded videos of HRI experiment. The tool was created using a new embedded dataset consisting of 300 smiling faces. By using 5-fold cross validation technique to train and validate, the image recognition accuracy obtained was 90%.

Conclusively, the smile recognition tool successfully recognised smile occurrences in the HRI experiment involving patients with post-stroke depression and animal-robot PARO. The results are compared with the psychological tools. Statistically, the smile detection tool has significant results with the psychological tools. This suggests that the results produced by HRI smile detection tools can be used to evaluate response during the human-robot interaction.

#### ACKNOWLEDGMENT

The authors gratefully acknowledge the Ministry of Higher Education Malaysia, Universiti Teknikal Malaysia Melaka, University of Jeddah and SOCSO Tun Razak Rehabilitation Center (TRRC) Melaka for their support. This project is funded under the Fundamental Research Grants Scheme (FRGS) [FRGS/1/2016/SKK06/FKP-AMC/ F00321].

#### REFERENCES

- [1] C. Breazeal, "Toward sociable robots," in *Robotics and Autonomous Systems*, 2003, vol. 42, no. 3–4, pp. 167–175.
- [2] S. Shamsuddin, N. A. Malik, H. Hashim, H. Yusof, F. A. Hanapiah, and S. Mohamed, "CCIS 376 - Robots as Adjunct Therapy: Reflections and Suggestions in Rehabilitation for People with Cognitive Impairments," *Ccis*, vol. 376, pp. 390–404, 2013.
- [3] C. Bühler, "Robotics for rehabilitation - A European (?) perspective," *Robotica*, vol. 16, no. 5, pp. 487–490, 1998.
- [4] S. Šabanovic, C. C. Bennett, W. L. Chang, and L. Huber, "PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia," in *IEEE International Conference on Rehabilitation Robotics*, 2013.
- [5] R. Aminuddin, A. Sharkey, and L. Levita, "Interaction with the paro robot may reduce psychophysiological stress responses," *ACM/IEEE Int. Conf. Human-Robot Interact.*, vol. 2016–April, pp. 593–594, 2016.
- [6] A. F. G. Leentjens, F. R. J. Verhey, R. Lousberg, H. Spitsbergen, and F. W. Wilmink, "The validity of the Hamilton and Montgomery-Åsberg depression rating scales as screening and diagnostic tools for depression in Parkinson's disease," *Int. J. Geriatr. Psychiatry*, vol. 15, no. 7, pp. 644–649, 2000.
- [7] C. Chang et al., "Multi-level Smile Intensity Measuring Based on Mouth-Corner Features for Happiness Detection," pp. 181–184, 2014.
- [8] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of database systems*, Springer, 2009, pp. 532–538.
- [9] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Toward practical smile detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2106–2111, 2009.
- [10] J. Michael, M. K. Lyons, and J. Gyoba, "Japanese female facial expressions (JAFFE)," *Database Digit. images*, 1997.
- [11] J. T. Mundt, N. F. Glenn, K. T. Weber, T. S. Prather, L. W. Lass, and J. Pettingill, "Discrimination of hoary cress and determination of its detection limits via hyperspectral image processing and accuracy assessment techniques," *Remote Sens. Environ.*, 2005.