ORIGINAL

Reliability of Emotion Analysis from Human Facial Expressions Using Multi-task Cascaded Convolutional Neural Networks

Toshiya Akiyama[1], Allan Paulo L. Blaquera[1,2], Leah Anne Christine Bollos[1], Gil P. Soriano[3], Hirokazu Ito[4], Ryuichi Tanioka[5], Hidehiro Umehara[6], Kyoko Osaka[7], and Tetsuya Tanioka[8]

[1]PhD Student, Graduate School of Health Sciences, Tokushima University, Tokushima, Japan, [2]School of Nursing and Allied Health Sciences, St. Paul University Philippines, Tuguegarao City, Philip-pines, [3]Department of Nursing, College of Allied Health, National University, Manila, Philippines, [4]Associate Professor, Department of Nursing Art and Sciences, Tokushima University, Tokushima, Japan, [5]Lecturer, Department of Physical Therapy, Hiroshima Cosmopolitan University, Hiroshima, Japan, [6]Department of Psychiatry, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan, [7]Professor, Department of Psychiatric Nursing, Nursing Course of Kochi Medical School, Kochi University, Kochi, Japan, [8]Professor, Department of Nursing Outcomes Management, Institute of Biomedical Sciences, Tokushima University, Tokushima, Japan

Abstract : Life support robots in care settings must be able to read a person's emotions from facial expressions to achieve empathic communication. This study aims to determine the degree of agreement between Multi-task Cascaded Convolutional Neural Networks (MTCNN) results and human subjective emotion analysis as a function to be installed in this type of robot. Forty university students talked with PALRO robot for 10 minutes. Thirteen area of interest videos were used to assess the validity identified by MTCNN was facial expression was happy or combination of happy and other emotions. Twenty university students and 20 medical professionals identified which of the 7 emotions (angry, disgust, fear, happy, sad, surprise, neutral) were present. Fleiss' kappa coefficient was calculated. Kappa coefficients of the emotion analysis for seven emotions ranged from 0.21 to 0.28. Kappa coefficient for "Happy" was the highest (0.52 to 0.57) with moderate agreement. Among female university students, only "Surprise" had a moderate agreement with Fleiss' kappa coefficient of 0.48. MTCNN emotion analysis and human emotion analysis were in moderate agreement for the identification of "Happy" emotions. The comparison of the agreement between the results of emotion analysis from facial expressions using non-contact MTCNN and subjective human facial expression analysis suggested that the use of MTCNN may be effective in understanding subjects' happy feelings. J. Med. Invest. 72 : 93-101, February, 2025

Keywords **:** *Inter-rater reliability, Facial expression analysis, Multi-task Cascaded Convolutional Neural Networks, Robotics*

INTRODUCTION

Japan's population is aging rapidly, and it is projected that by 2050, the middle age in Japan will be 53.6 years old, and the percentage of the population over 65 years old will be over 37.5% (1). This has led to an increased demand for long-term care services, such as facilities for the older population and long-term care facilities (2, 3). Therefore, medical professionals should improve the quality of care for the older population in Japan's hyper-aged society (4).

Japan's Ministry of Health, Labour and Welfare's 2022 report states that an additional 320,000 care professionals will be needed by 2025, compared with the number of care professionals in 2019 (5). However, due to declining birthrate and population aging, the number of people aged 20 to 64 is projected to decrease by 14 million human resources over the next 20 years (6). In developed countries, such as Japan, where the birthrate is declining and the population is aging, there is a shortage of nursing staff in chronic care wards (7). Thus, communication robots are being used to maintain social activities, promote communication, prevent dementia, and improve the effectiveness of treatment for older adults who live alone and otherwise have little social participation in Japan (7). In the United States, Korea, Australia, and other countries, artificial intelligence (AI)-based social support robots that observe and manage the health of the older population are also attracting attention (8).

A meta-analysis of the effect of social support robots using artificial intelligence on the improvement of cognitive function in the older population aged 65 years and older from 2010 to 2021 reported significant improvements in cognitive function (9). In addition, it has been reported that social support robots not only support the older population but also the caregivers who assist the older population (10).

In 2016, Japan's Cabinet Office launched the AI Technology Strategy Council, which cited Society 5.0, the 5th Science and Technology Basic Plan for the future of Japanese society using AI, as an important technological foundation (11, 12). In Society 5.0, research on human-machine interaction will be important to establish a harmonious relationship in which humans and social support robots work together toward a common goal in physical space (13).

Interpersonal trust refers to a psychological state that includes the intention to accept vulnerability to the other person's behavior in the expectation that the other person will take certain actions that are important to them (14). When communicating with patients, interpersonal relationships stagnate without trust, and patients and their families may not be able to provide all the care they need (15). A social support robot that act in concert with humans, it must be trusted by humans (16). Trusted relationships are also important for improving interpersonal relationships with patients (17).

Received for publication October 19, 2024 ; accepted November 10, 2024.

Address correspondence and reprint requests to Tetsuya Tanioka, Professor, Department of Nursing, Institute of Biomedical Sciences, Tokushima University, Graduate School, Tokushima, 770-8509, Japan. E-mail : tanioka.tetsuya@tokushima-u.ac.jp

Robot characteristics, such as trustworthiness, behavior, and transparency have been shown to influence peoples' level of trust in social support robots (18). Empathy refers to the ability to put oneself in another person's "heart position" and understand their emotions and feelings (19). Translating this to robots, trust between people and social support robots requires the ability of social support robots to infer human emotions (artificial empathy) (20-23).

Artificial empathy by robots refers to the ability of computer systems to understand and respond to human thoughts, feelings, and emotions (24). Robot empathy with humans is also said to include the robot's ability and process to recognize human emotional states, thoughts, and situations and to generate emotional or cognitive responses (25).

When people estimate the emotions of others, they do so primarily based on facial expressions and voice information (26). The importance of emotion estimation results based on facial expressions and voice varies by culture and age (27, 28), and individual physical cues, such as facial expressions, speech, posture, and gestures, are often invalidated by social masking, in which people consciously or unconsciously hide their true emotions (29). Therefore, physiological signals, such as heart rate variability, electroencephalography signals, and eye movements can be used to estimate emotions accurately and objectively (30, 31).

However, it is impractical for social assistance robots to wear contact-type measuring devices on humans in everyday life to infer human emotions. In public places, physical cues, such as facial expressions, conversation, posture, and gestures have proven to be much more useful and effective (32-34). Over the years, advances in deep learning have improved the accuracy of facial expression analysis (35). Artificial intelligence algorithms, such as deep learning and reinforcement learning have evolved, and certain algorithms, such as convolutional neural networks and recurrent neural networks have attracted attention owing to their ability to analyze images, voice, and even video (36).

Facial expression recognition systems are an emerging field of computer vision research that focuses on developing technologies that can sense, understand, and respond to human emotions (37). Automated computer vision technology has made it easier and cheaper to analyze large amounts of data and to consistently quantify facial features due to advances in affective computing. Improved techniques that leverage advances in temporal and spatial granularity have also allowed computer vision-based analysis to provide researchers with a deeper understanding of the phenomena of their subjects (38).

In vision-based facial expression recognition, independent feature extraction and classification are the main concerns. The selection and classification of salient features requires expertise, is time-consuming, and is error-prone (39). Several studies have used convolutional neural networks to address the limitations of traditional vision-based facial expression recognition (40) and have showed improved performance (41, 42). Zhang *et al*. (43) used Multi-task Cascaded Convolutional Neural Networks (MTCNN) to jointly perform face detection and alignment by coarsely and finely predicting the positions of faces and landmarks. Using MTCNN improved the accuracy of real-time face detection, enabling rapid face detection and showing the best comprehensive facial recognition performance (44). Human communication relies mostly on nonverbal signals expressed through body language. Facial expressions, in particular, convey emotional information that allows people involved in social interactions to determine each other's emotional states and adjust their behavior appropriately (45).

The significance of this study is to address whether a social support robot can analyze human emotions using facial expression analysis based on a non-contact MTCNN in the same way that a person can guess another person's emotions without using a contact-type device that receives biological signals, such as heart rate variability and cerebral blood flow. This enables the social support robot to appropriately infer emotions from the facial expressions of others and speak to them empathetically.

This study aims to determine the degree of inter-rater agreement between the MTCNN facial expression analysis results and human subjective facial expression analysis as a function to be installed in this type of robot.

MATERIALS AND METHODS

Theoretical framework

Peplau (46) described that the nurse-patient relationship as "the primary human contact that is central fundamental to providing nursing care." The Interpersonal Relations theory (46) relies on participant observation. One type of participant observation is "empathic linkages," which is the "ability to feel in oneself the emotions experienced by another person in the same situation" (46). The data from these patient observations result in the nurse's understanding of the patient, which will serve as a basis for the nurse to decide what nursing care the patient requires.

When a person shows empathic understanding to another person, he or she judges the other person's emotions from the other person's voice, facial expressions, actions (47, 48), and conveys that he or she empathizes with the other person (48, 49). For a robot to show empathy to a human, it must be able to grasp the other person's emotions accurately.

Visual modality allows the robot to perceive and interpret visual cues, such as facial expressions, gestures, body language, and gaze direction. Robots that employ visual modalities can perceive their environment using cameras and process visual information using computer vision algorithms. This modality is used in applications, such as robot navigation, surveillance, and human-robot interaction (50). This study focused on whether the MTCNN can grasp the other person's emotion via facial recognition using a visual modality as a required function for robots.

Study Subjects

To ensure that the results of the experiment were not influenced by the data of the subjects, 40 university students in their 20s were included in the study, whose expressions varied according to age, gender, and nationality. Note that since the subjects were conversing with a communication robot, severe communication disorders were used as an exclusion criterion.

Data Collection Period

The study was conducted from June 30, 2023, to August 3, 2023.

Social Support Robot to be used in this study

PALRO (a communication robot 40 cm tall and weighing 1.8 kg) from Fujisoft, Ltd. was used as a social support robot. It can perform smooth conversation and is suitable for communicating with people ; therefore, it can perform daily conversation, recreation, and health exercise guidance. In addition, linking with PALRO's dedicated application makes the intended conversations and other activities possible. In Japan, PALRO has been selected as a communication robot for investigation by the Ministry of Economy, Trade and Industry's "Project to Promote the Development and Introduction of Robotic Nursing Care Equipment" in recognition of the evidence from various demonstration tests using PALRO (51).

Clinical Experimental Methods

The experimental site was at the Department of Health Sciences, University of Tokushima. The communication mode within PALRO was used for intentional conversation with the subject. This allows the operator to remotely control the words and movements of the communication robot. A scripted conversation can be added remotely using a tablet or keyboard. The length of each subject's conversation with the communication robot is approximately 10 minutes. As shown in Figure 1, the interaction between PALRO and one subject was recorded using a single digital video camera. The operator typed the content of the communication robot's conversation with the subject. One examiner contacted the operator assistant to coordinate the end time and record important events during the conversation and radio clock time in a field notebook (52, 53). The subject is situated against a wall and in a way where no one else can be seen in the background. The distance between the subject and the communication robot was 80 cm. To recognize facial expressions, the subject's eyes, mouth, and nose were made visible.

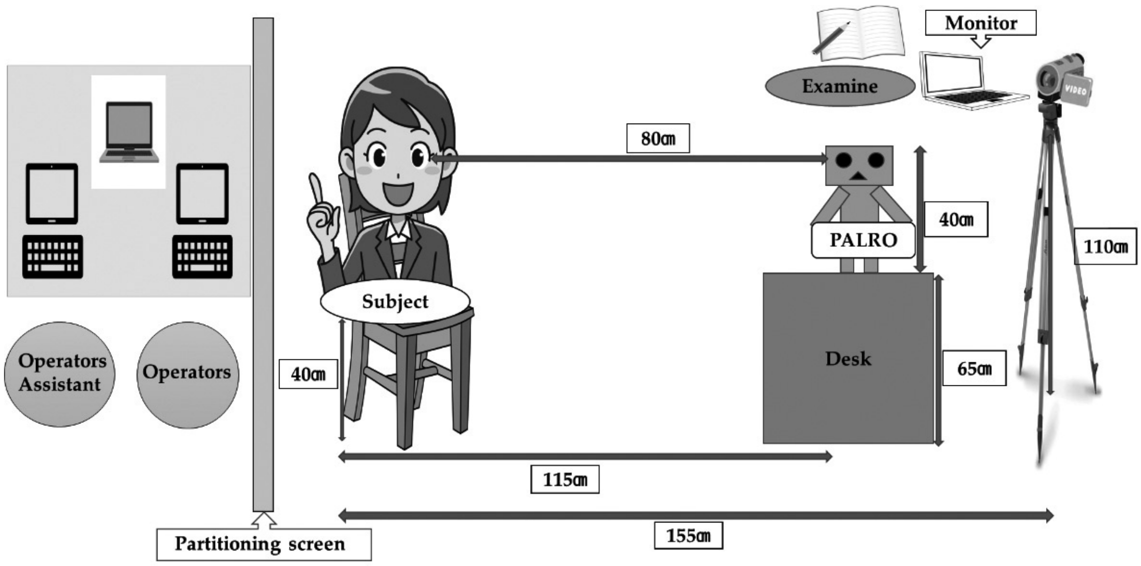

Figure 1. Experimental environment

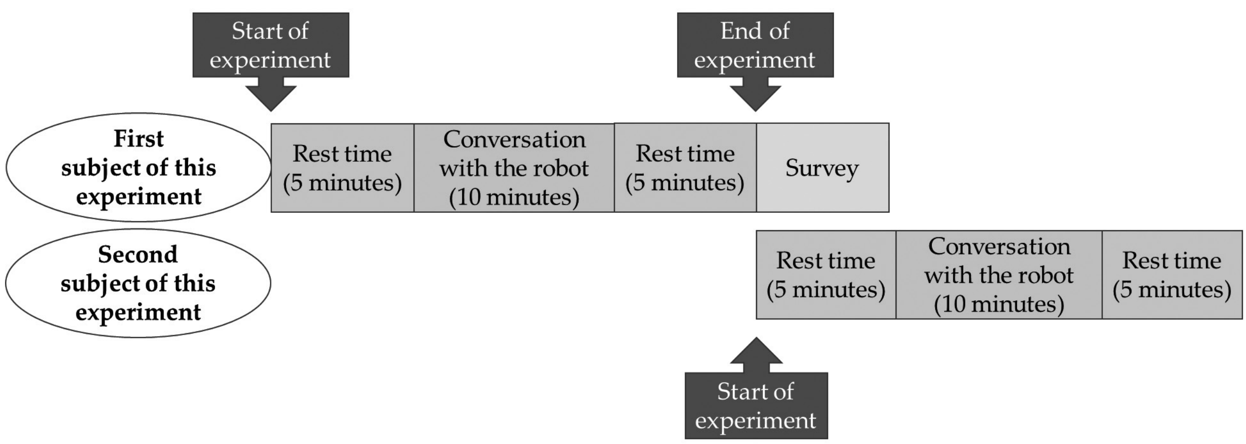

Figure 2. Time schedule of the experiment

The process of selecting data for analysis

Of these 40 subjects, 28 were selected who had no missing data. Next, we selected 16 students whose data indicated in the questionnaire that they were happy to talk with the communication robot. Finally, the MTCNN facial analysis results of 15 subjects who were "happy" with at least one video frame were selected for analysis.

Analysis of facial expressions using Multi-task Cascaded Convolutional Neural Networks

The objective method employed the MTCNN which included in the OpenCV library was used. The MTCNN has been used the facial expression recognition class (54), an open-source library for Python. The facial expression recognition class accepts the parameter "mtcnn" as an argument, but the default parameter is "mtcnn = False". The facial expressions to be analyzed are classified into 7 categories : angry, disgust, fear, happy, sad, surprise, and neutral, using a vision-based facial expression recognition classification model, and analyzed using the MTCNN, which is a deep learning-based method. MTCNN performs highly accurate face detection. It uses convolutional neural networks (P-Net, R-Net, and O-Net) with three different resolution images as input to perform the three tasks (face classification, bounding box regression, and face feature point detection) (55). A combination of MTCNN-based face detection and Deep Convolutional Neural Network based facial expression recognition models were used to determine the subject's emotional state.

Method for video processing and extraction of area of interest from video

The conversation videos selected for facial expression analysis were from the subjects who mentioned in their respective post-experiment questionnaires that they felt happy conversing with PALRO.

The video editing software Shotcut (Dun & Bradstreet, Inc.) was used for facial expression extraction. Shotcut removes all unnecessary parts of the subject's face where different facial expressions were identified. Mosaic processing was applied to the areas unrelated to the facial expression analysis by artificial intelligence. Video size was adjusted to a frame width of 640, frame height of 480, and frame rate of 29.97 frames per second. To allow minute changes in facial expressions to be analyzed, video frames judged to be happy data were extracted using the editing software and were used for evaluation.

The analysis results were displayed relative to the seven facial expressions ; thus, the sum of all emotions was 100%. The analysis output of the video was 30 frames per second. The 337 MTCNN facial analysis results show happy at least one video frame were selected. Within these results, parts where happy was present for more than one frame in a row for more than 6 seconds were selected.

Finally, 13 area of interest in which the subject's facial expression was "happy" or the sum of "happy" and other emotions were identified in 162 frames or more (90% or more of 180 frames in 6 seconds), which were selected for the validity assessment. Table 1 shows 13 area of interest videos to assess the validity identified by MTCNN was facial expression was happy or combination of happy and other emotions.

Raters and validity assessment method

There were 40 raters comprising 10 male and 10 female university students and 10 male and 10 female healthcare professionals. The university student raters consisted of newly enrolled freshmen to reduce the possibility of differential learning experience bias. Healthcare professionals with more than 10 years of experience were selected because of their experience in interacting with people and their ability to understand facial expressions (56, 57). These raters analyzed 13 extracted target videos in silence and selected their evaluation result based on seven emotions, the same as the facial expressions analyzed by MTCNN.

Statistical Processing Method

The statistical procedure for this study was to determine the minimum, median, and maximum values for the subjects and raters. Agreement between raters was measured using the overall agreement rate and the Fleiss' kappa coefficient. In this study, a significance probability (p-value) of less than 0.05 is considered statistically significant.

Table 1. Emotional percentages of the 13 happy focus videos used in the validity assessment identified by MTCNN

Area of interest	Percent of happy	Percent of other emotions
1	**100%**	0%
2	**100%**	0%
3	**49%**	**Disgust : 50%**, Angry : 1%
4	**100%**	0%
5	**34%**	**Neutral : 42%, Disgust : 24%**
6	**100%**	0%
7	**41%**	**Angry : 55%**, Disgust : 1%, Sad : 3%
8	**98%**	Disgust : 2%
9	**100%**	0%
10	**97%**	Angry : 1%, Disgust : 1%, Sad : 1%
11	**100%**	0%
12	**53%**	**Neutral : 44%**, Angry : 3%
13	**100%**	0%

RESULTS

Table 2 shows the demographic characteristics of the subjects and raters. All subjects in this study were female, with a minimum age of 18 years, a maximum age of 22 years, and a median age of 20 years. Raters comprised university students and healthcare professionals from both genders. All university student raters were 18 years old. The minimum, maximum, and median ages of male healthcare professional raters were 34, 54, and 37.5 years, respectively, and the minimum, maximum, and median years of professional employment were 10, 31, and 14.5 years, respectively. The minimum, maximum, and median ages of female healthcare professionals were 34, 47, and 40 years, respectively, and the minimum, maximum, and median years of professional employment were 11, 25, and 17.5 years, respectively.

The professional occupations of the females were 7 nurses, 1 pharmacist, 1 physical therapist, and 1 dietitian, and the professional occupations of the males were 7 nurses, 2 pharmacists, and 1 laboratory technician.

Table 3 shows the agreement between the MTCNN-rated analysis of the seven different emotions and the raters' evaluations based on the same facial expressions.

The overall Fleiss' kappa coefficient for the seven emotions judged by humans was 0.21-0.28, which is interpreted as a slight agreement. The Fleiss' kappa coefficient for the human judgment of the emotion happy was 0.52-0.57, which indicates a moderate agreement. Fleiss' kappa coefficient for the emotion of surprise was 0.48 (moderate agreement), was seen among the female university student raters.

University student males were Happy ($p<0.01$), Surprise ($p=0.01$), Fear ($p=0.01$), and Neutral ($p<0.01$). University student females were Happy ($p<0.01$), Sad ($p<0.01$), Disgust ($p<0.01$), Surprise ($p<0.01$), and Neutral ($p<0.01$). Healthcare professional males were Happy ($p<0.01$), Sad ($p=0.01$), and Neutral ($p<0.01$). Healthcare professional females were Happy ($p<0.01$), Disgust ($p<0.01$), and Neutral ($p=0.02$). The indicated locations showed significant differences at $p<0.05$, but the locations with low Fleiss' kappa coefficient resulted in low correlations.

DISCUSSION

This study sought to determine the degree of inter-rater agreement between human raters and MTCNN regarding facial expressions of emotions. There was a slight to moderate agreement across the examined emotions, with Happy having the highest levels of agreement among the human raters and the MTCNN (58).

The visual perception of emotions can vary across age, sex, social background, and culture. Tu *et al.* (59) summarize several studies that reveal that Western and East Asian participants have different viewing strategies when processing facial stimuli. Westerners tend to focus on salient features and use those attributes to categorize what they are analyzing (60). On the other hand, East Asians tend to analyze more holistically, and this can lead to overlapping judgments about certain emotions (60). The degree of inter-rater agreement may be attributed to the difference in culture between the raters who are of Japanese descent and MTCNN, which is trained on Western-based examples (55, 61).

In addition, since it was determined that a person with suffi-

Table 2. Demographic characteristics of subjects and raters

Target Subjects		Gender	Age (Years old)			Years of professional experience (Years)		
			Min.	Median	Max.	Min.	Median	Max.
Subjects		Female (N = 15)	18	20	22			
Raters	University students	Male (N = 10)	18	18	18			
		Female (N = 10)	18	18	18			
	Healthcare professionals	Male (N = 10)	34	37.5	54	10	14.5	31
		Female (N = 10)	34	40	47	11	17.5	25

Min. = minimum, Max. = maximum.

Table 3. Degree of agreement between MTCNN and human subjective emotion analysis

			Happy	Angry	Sad	Disgust	Surprise	Fear	Neutral	Total κ	Total p
University students (N = 20)	Male (N = 10)	κ	0.52	-0.01	0.00	0.03	0.11	0.10	0.16	0.22	0.00
		p	<0.01	1.15	0.95	0.44	0.01	0.01	<0.01		
	Female (N = 10)	κ	0.57	-0.02	0.18	0.17	0.48	0.08	0.13	0.28	0.00
		p	<0.01	1.29	<0.01	<0.01	<0.01	0.06	<0.01		
Healthcare professional (N = 20)	Male (N = 10)	κ	0.57	-0.02	0.10	0.06	0.06	0.01	0.17	0.24	0.00
		p	<0.01	1.29	0.01	0.15	0.14	0.81	<0.01		
	Female (N = 10)	κ	0.53	0.02	0.08	0.24	0.06	-0.02	0.09	0.21	0.00
		p	<0.01	0.57	0.05	<0.01	0.15	1.29	0.02		

κ : Fleiss' kappa coefficient p : P value cient years of clinical experience would have the ability to adequately infer the emotions of others, persons with 10 to 31 years of clinical experience were selected as evaluators (48, 49). For the student evaluator, we selected a first-year university student who did not know the subject because having someone who knew the subject's usual condition could cause positivity bias.

Inter-rater reliability was measured using the Fleiss kappa coefficient (62). Overall agreement for the seven emotions in this study was good (Fleiss' K = 0.21-0.28), and moderate agreement was obtained for "Happy" (Fleiss' K = 0.52-0.57). This result is congruent with the findings of Rutter *et al*. (63), who found that the ability to recognize "Happy" emotion remained preserved throughout the lifespan.

Only female students showed moderate agreement (Fleiss' K= 0.48) for "Surprised." Thirteen area of interest videos to assess the validity identified by MTCNN was facial expression was happy (ranged from 34% to 100%) or combination of happy and other emotions (mainly, disgust, neutral, and angry).

This follows a similar finding in a study where young females generally perceived emotions of higher intensities (fear, angry, or happy) more than young males (63). This also corresponds to the fact that young adults have a higher ability to recognize facial expressions than older adults (64). It was reported that fearful, neutral, and disgusted faces, with surprised faces being most likely classified as happy faces (65). In our study, thus, it was considered that the facial expressions, which female students evaluated as surprising were a disgust, angry, and neutral.

Cîrneanu *et al*. (66) noted that the technology to accurately track users' feelings accurately is not sufficiently mature and still remains a challenge. If emotion analysis based on facial expressions using non-contact MTCNN becomes feasible, humanoid robots could be used in healthcare.

The number of papers regarding the facial expression recognition is increasing, however, there is no uniformity in which of the existing data sets is used (67). That is, existing publicly available data sets are insufficient for effective facial expression recognition and are not diverse enough. Effective solutions to these problems are needed, including data expansion, combining multiple datasets, modifying existing data, and creating new datasets (68). Despite the availability of large annotated training datasets, challenges remain due to image size variation and unbalanced categories. Inherent biases can lead to suboptimal performance in certain classes and pose a risk of overtraining (69). It is this risk that creates misclassification, the most influential issue in facial expression recognition datasets (70). Certain facial expressions, such as fear and surprise, disgust and anger, contempt and sadness, have only slight differences. Confusion is possible when multiple emotions are simultaneously felt, such as laughing while shedding tears (71). These factors have been shown to result in mislabeling and placing facial images in the wrong directory (70). Therefore, due to fewer face landmarks and their intensity for deep learning models, performance improvement for facial expression recognition still needs to be improved for accurately predicting facial emotion recognition. Zoning-based face expression recognition method was proposed that to locate more face landmarks to perceive deep face emotions indemnity through zoning all facial zones (left eye, right eye, nose, mouth, forehead) (72).

Nursing is a discipline that develops knowledge of the relationship between compassion and health, healing, and well-being (73). Humanoid robots need artificial empathy to "understand" the feelings and thoughts of others and respond accordingly (74). This requires remembering experiences (75), processing language, maintaining conversations (69), and reading emotions from facial expressions (76). For future research, the use of humanoid robots in nursing requires not only the technical dexterity that humanoid robots excel at but also the communication skills to interact with patients, including the ability to listen to patients with compassion, understand people who are getting better, and make ethical decisions accurately and appropriately (76).

IMPLICATIONS

The results of this study suggest that MTCNN-based emotion analysis can be applied to humanoid robots in healthcare settings. However, to convey empathy, these robots would require advanced capabilities in emotion recognition and conversation. The use of MTCNN also allows for non-contact and continuous monitoring of patients' emotional states. This is useful in settings where patients are unable to communicate, such as in intensive care units or for patients with speech impairments. MTCNN can also provide real-time feedback on patients' emotional well-being, enabling timely interventions and improving overall patient care.

LIMITATIONS OF THIS STUDY

Since the subjects in this study were limited to university students, the accuracy of the MTCNN in determining emotions from non-contact facial expressions may also vary depending on the gender, age, and medical history of the subjects. This study conducted an emotion assessment without speech, it is possible that the recognition of emotions from facial expressions may have been ambiguous compared with the recognition of muted facial expressions associated with nonverbal vocalizations and verbal information. The small amount of data used in this study limits its generalizability.

This is because the facial expression recognition dataset (FER2013) used in this study was trained from Westerners' facial photos. Therefore, if we had used the Oriental-based facial expression recognition dataset, although such a database does not actually exist, the evaluation results of facial expressions might have been different, and even more accurate evaluation results could have been obtained. To confirm the validity and reliability of the MTCNN for practical use, studies are needed in the older populations and in people with medical conditions and different cultural backgrounds that make reading facial expressions difficult.

CONCLUSION

The comparison of the agreement between the results of emotion analysis from facial expressions using non-contact MTCNN and subjective human facial expression analysis suggested that the use of MTCNN may be effective in understanding subjects' happy feelings. In human communication, we gradually become aware of the emotions of others through facial expressions. In the human-robot interaction, it can be gradually understanding each other. By programing in this way, it is expected that the accuracy of facial emotion estimation by MTCNN will also be improved.

CONFLICT OF INTEREST

All authors declare no actual or potential conflicts of interest associated with this study.

CONTRIBUTIONS

Author Contributions : Conceptualization, T.A., and T.T. ; methodology, H.U., K.O., and T.T. ; software, R.T. ; validation, H.I., and A.B. ; formal analysis, T.A., R.T., H.I., K.O., and A.B. ; investigation, K.O., T.A., R.T., L.B., and T.T. ; writing—original draft preparation, A.B., G.S., T.A., R.T., L.B. and T.T. ; writing—review and editing : K.O., R.T., H.U., H.I., T.T. and A.B. ; project administration : T.T. All authors have read and agreed to the publication of the finale version of the manuscript.

FUNDING

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT

The study was conducted in accordance with the Declaration of Helsinki, and was approved by the Ethics Committee of the Tokushima University Hospital (#3046).

INFORMED CONSENT STATEMENT

The details of the experiment, the experimental tasks, and the handling of personal information were explained along with a written explanation and consent form for experimental cooperation that summarized the rewards for cooperation. All participants provided written informed consent before participating in the study.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors on request.

ACKNOWLEDGMENTS

The authors would like to express their gratitude and deep appreciation to the participants.

REFERENCES

1. Nakatani H : Ageing and shrinking population : The looming demographic challenges of super-aged and super-low fertility society starting from Asia. Global health & medicine 5(5) : 257-63, 2023
2. O.E.C.D. Health at a Glance 2021 : OECD Indicators. OECD Publishing, Paris, 2021
3. Organization WH : World report on ageing and health. World Health Organization, 2015
4. Zhao D, Sun X, Shan B : Research status of elderly-care robots and safe human-robot interaction methods. Front Neurosci 17(1291682) : 2023
5. M.H.L.W. : Actions toward ensuring human resources for care [Internet]. 2021. Available from : https://www.mhlw.go.jp/stf/newpage_02977.html
6. Health M. Labour and Welfare White Paper 2022. Ministry of Health, Labour and Welfare. 000988388(pdf)
7. You W, Donnelly F : Greater nurse density correlates to higher level of population ageing globally, but is more prominent in developed countries. PLoS One 18(9) : 2023
8. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP : Scoping Review on the Use of Socially Assistive Robot Technology in Elderly Care. BMJ Open 8 : 018815, 2018
9. Lee H, Chung MA, Kim H, Nam EW : The Effect of Cognitive Function Health Care Using Artificial Intelligence Robots for Older Adults : Systematic Review and Meta-analysis [published correction appears in JMIR Aging. JMIR Aging 5(2) : e38896, 2022
10. Sawik B, Tobis S, Baum E : Robots for Elderly Care : Review, Multi-Criteria Optimization Model and Qualitative Case Study. Healthcare (Basel) 11(9) : 2023
11. Office C : Report on the 5th Science and Technology Basic Plan Council for Science [Internet]. Technology and Innovation ; 2015. Available from : https://www8.cao.go.jp/cstp/kihonkeikaku/5basicplan_en.pdf
12. The Government of Japan : Realizing Society 5.0[internet] JapanGov [Internet]. 2020. Available from : https://www.japan.go.jp/abenomics/_userdata/abenomics/pdf/society_5.0.pdf
13. Mourtzis D, Angelopoulos J, Panopoulos N : The Future of the Human–Machine Interface (HMI) in Society 5.0. Future Internet [Internet]. 15(5) : 2023. Available from : https://doi.org/10.3390/fi15050162
14. Six FE : Building interpersonal trust within organizations : a relational signalling perspective. J Manage Governance 11 : 285-309, 2007
15. Sisk B, Baker JN : A Model of Interpersonal Trust, Credibility, and Relationship Maintenance. Pediatrics 144(6) : 2019
16. Brave S, Nass C, Hutchinson K : Computers that care : investigating the effects of orientation of emotion exhibited by an embodied computer agent. International Journal of Human-Computer Studies 62(2) : 161-78, 2005
17. Heyns M, Rothmann S : Dimensionality of trust : an analysis of the relations between propensity, trustworthiness and trust. SA J Industrial Psychol 41 : 1263, 2015
18. Hancock PA, Billings DR, Schaefer KE, Chen JYC, Visser EJ, Parasuraman R : A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. Human Factors 53(5) : 517-27, 2011
19. Decety J, Lamm C : Human empathy through the lens of social neuroscience. ScientificWorldJournal 6 : 1146-63, 2006
20. Broadbent E : Interactions With Robots : The Truths We Reveal About Ourselves. Annu Rev Psychol 68 : 627-52, 2017
21. Davis MH : Empathy : a social psychological approach. Westview Press, Boulder, Colo., 1996, 260 p. (Social psychology series)
22. Sorbello R, Chella A, Giardina M, Nishio S, Ishiguro H : An Architecture for Telenoid Robot as Empathic Conversational Android Companion for Elderly People. In : Menegatti E, Michael N, Berns K, Yamaguchi H, editors, Intelligent Autonomous Systems 13 [Internet]. Cham : Springer International Publishing, 2016 [cited 2024 Jun 30], p. 939-53. (Advances in Intelligent Systems and Computing ; vol. 302). Available from : https://link.springer.com/10.1007/978-3-319-08338-4_68
23. Tanioka T, Yokotani T, Tanioka R, Betriana F, Matsumoto K, Locsin R, Zhao Y, Osaka K, Miyagawa M, Schoenhofer S : Development Issues of Healthcare Robots : Compassionate Communication for Older Adults with Dementia. IJERPH 18 : 4538, 2021
24. Yalçın ÖN, DiPaola S : Modeling empathy : building a link between affective and cognitive processes. Artif Intell Rev 53(4) : 2983-3006, 2020
25. Park S, Whang M : Empathy in Human-Robot

Interaction : Designing for Social Robots. Int J Environ Res Public Health 19(3) : 2022
26. Karle KN, Ethofer T, Jacob H, Brück C, Erb M, Lotze M, Nizielski S, Schütz A, Wildgruber D, Kreifelts B : Neurobiological Correlates of Emotional Intelligence in Voice and Face Perception Networks. Soc Cogn Affect Neurosci : 233-244, 2018
27. Grossard C, Chaby L, Hun S, Pellerin H, Bourgeois J, Dapogny A, Ding H, Serret S, Foulon P, Chetouani M, Chen L, Bailly K, Grynszpan O, Cohen D : Children Facial Expression Production : Influence of Age, Gender, Emotion Subtype, Elicitation Condition and Culture. Front Psychol 2018
28. Kikutani M : Influence of social anxiety on recognition memory for happy and angry faces : Comparison between own- and other-race faces. Quarterly Journal of Experimental Psychology 71(4) : 870-8, 2018
29. Houssein EH, Hammad A, Ali AA : Human emotion recognition from EEG-based brain–computer interface using machine learning : a comprehensive review. Neural Comput & Applic 34(15) : 12527-57, 2022
30. Samadiani N, Huang G, Cai B, Luo W, Chi CHC, Xiang Y, He J : A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data. Sensors 19 : 2019
31. Akiyam T, Matsumoto K, Osaka K, Tanioka R, Betriana F, Zhao Y, Kai Y, Miyagawa M, Yasuhara Y, Ito H, Soriano GP, Tanioka T : Comparison of Subjective Facial Emotion Recognition and “Facial Emotion Recognition Based on Multi-Task Cascaded Convolutional Network Face Detection” between Patients with Schizophrenia and Healthy Participants. Healthcare 10 : 2022
32. Shi Y, Zheng X, Li T : Unconscious Emotion Recognition based on Multi-scale Sample Entropy. In : 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) [Internet]. IEEE, Madrid, Spain, 2018 [cited 2024 Jun 30], p. 1221–6. Available from : https://ieeexplore.ieee.org/document/8621185/
33. Thomson DMH, Coates T : Are unconscious emotions important in product assessment? How can we access them? Food Quality and Preference 92 : 104123, 2021
34. Kaklauskas A, Abraham A, Ubarte I : A Review of AI Cloud and Edge Sensors, Methods, and Applications for the Recognition of Emotional, Affective and Physiological States. Sensors (Basel) 22(20) : 7824, 2022
35. Hung LP, Lin CC : A multiple warning and smart monitoring system using wearable devices for home care. International Journal of Human-Computer Studies 136 : 102381, 2020
36. Pappas IO, Mikalef P, Dwivedi YK, Jaccheri L, Krogstie J : Responsible Digital Transformation for a Sustainable Society. Inf Syst Front 25(3) : 945-53, 2023
37. Canal FZ, Müller TR, Matias JC, Scotton GG, De Sa Junior AR, Pozzebon E, Sobieranski AC : A Survey on Facial Emotion Recognition Techniques : A State-of-the-Art Literature Review. Information Sciences 582 : 593-617, 2022
38. Matsugu M, Mori K, Mitari Y, Kaneda Y : Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Networks 16(5-6) : 555-9, 2003
39. Alsharekh MF : Facial Emotion Recognition in Verbal Communication Based on Deep Learning. Sensors (Basel) 22(16) : 6105, 2022
40. Zhao S, Cai H, Liu H, Zhang J, Chen S : Feature Selection Mechanism in CNNs for Facial Expression Recognition. In : British Machine Vision Conference 2018, BMVC 2018. Newcastle, UK, September 3-6, 2018 [Internet]. 2018, p. 317. Available from : http://bmvc2018.org/contents/workshops/iahfar2018/0011.pdf
41. Lasri I, Solh AR, Belkacemi ME : Facial Emotion Recognition of Students Using Convolutional Neural Network. In : 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS) Marrakech, Morocco, 2019, p. 1-6
42. Lek JXY, Teo J : Academic Emotion Classification Using FER : A Systematic Review. E. Domoff S, editor. Human Behavior and Emerging Technologies : 1-27, 2023
43. Zhang K, Zhang Z, Li Z, Qiao Y : Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE Signal Process Lett 23(10) : 1499-503, 2016
44. Ge H, Dai Y, Zhu Z, Wang B, School of Electronic Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China : Robust face recognition based on multi-task convolutional neural network. MBE 18(5) : 6638-51, 2021
45. Lazzeri N, Mazzei D, Ben Moussa M, Magnenat-Thalmann N, De Rossi D : The influence of dynamics and speech on understanding humanoid facial expressions. International Journal of Advanced Robotic Systems 15(4) : 172988141878315, 2018
46. Peplau HE : Peplau’s Theory of Interpersonal Relations. Nurs Sci Q 10(4) : 162-7, 1997
47. Keltner D, Sauter D, Tracy J, Cowen A : Emotional Expression : Advances in Basic Emotion Theory. Journal of nonverbal behavior 43(2) : 133-60, 2019
48. Lussier MT, Richard C : Communication tips. Feeling understood : expression of empathy during medical consultations. Canadian family physician Medecin de famille canadien 53(4) : 640-1, 2007
49. Cooke AN, Bazzini DG, Curtin LA, Emery LJ : Empathic Understanding : Benefits of Perspective-Taking and Facial Mimicry Instructions are Mediated by Self-Other Overlap. Motivation and emotion 42(3) : 446-57, 2018
50. Su H, Qi W, Chen J, Yang C, Sandoval J, Laribi MA : Recent advancements in multimodal human–robot interaction. Front Neurorobot 17 : 1084000, 2023
51. FUJISOFT Inc : Koureisya Fukushi Shisetumuke, Kaiwarobotto Saisenntan! PALRO. (In Japanese) Available from : https://palro.jp/en/preventive-care/nursing-home.html
52. Osaka K : Development of the model for the intermediary role of nurses in transactive relationships with healthcare robots. Int J Hum Caring 24 : 265-74, 2020
53. Osaka K, Sugimoto H, Tanioka T, Yasuhara Y, Locsin RC, Zhao Y, Okuda K, Saito K : Characteristics of a Transactive Phenomenon in Relationships among Older Adults with Dementia, Nurses as Intermediaries, and Communication Robot. Intell. Control Autom 8 : 111-25, 2017
54. Python Software Foundation : Facial expression recognition from images [Internet]. [cited 2024 Feb 1]. Available from : https://pypi.org/project/fer/
55. Schroff F, Kalenichenko D, Philbin J : FaceNet : A Unified Embedding for Face Recognition and Clustering. 2015 [cited 2024 Jul 15]. Available from : https://arxiv.org/abs/1503.03832
56. Lazerus T, Ingbretsen ZA, Stolier RM, Freeman JB, Cikara M : Positivity bias in judging ingroup members’ emotional expressions. Emotion 16(8) : 1117-25, 2016
57. Awad D, Emery NJ, Mareschal I : Role of facial familiarity and emotional expression intensity in ensemble emotion perception. Atten Percept Psychophys 85(6) : 1990-2003, 2023

58. Hartling L, Bond K, Harvey K, Santaguida PL, Viswanathan M, Dryden DM : Developing and Testing a Tool for the Classification of Study Designs in Systematic Reviews of Interventions and Exposures [Internet]. Agency for Healthcare Research and Quality (US), Rockville (MD), 2010 [cited 2024 Jul 15]. (AHRQ Methods for Effective Health Care). Available from : http://www.ncbi.nlm.nih.gov/books/NBK52670/
59. Tu YZ, Lin DW, Suzuki A, Goh JOS : East Asian Young and Older Adult Perceptions of Emotional Faces from an Age- and Sex-Fair East Asian Facial Expression Database. Front. Psychol 9 : 2358, 2018. doi : 10.3389/fpsyg.2018.02358.
60. Nisbett RE : The Geography of Thought : How Asians and Westerners Think Differently ... and Why ; 1. Free Press trade paperback ed., [Nachdr.]. ; Free Press, New York, 2004 ; ISBN 978-0-7432-5535-6.
61. Ekman P, Friesen WV (1978) : Facial Action Coding System (FACS) [Database record]. APA PsycTests. https://doi.org/10.1037/t27734-000
62. McHugh ML : Interrater reliability : the kappa statistic. Biochem Med : 276-82, 2012
63. Rutter LA, Dodell-Feder D, Vahia IV, Forester BP, Ressler KJ, Wilmer JB, Germine L : Emotion Sensitivity across the Lifespan : Mapping Clinical Risk Periods to Sensitivity to Facial Emotion Intensity. J Exp Psychol Gen 148 : 1993-2005, 2019
64. Hayes GS, McLennan SN, Henry JD, Phillips LH, Terrett G, Rendell PG, Pelly RM, Labuschagne I : Task Characteristics Influence Facial Emotion Recognition Age-Effects : A Meta-Analytic Review. Psychol Aging 35 : 295-315, 2020
65. Calvo MG, Nummenmaa L : Time course of discrimination between emotional facial expressions : the role of visual saliency. Vision Res 51(15) : 1751-9, 2011. doi : 10.1016/j.visres.2011.06.001.
66. Cîrneanu AL, Popescu D, Iordache D : New Trends in Emotion Recognition Using Image Analysis by Neural Networks, a Systematic Review. Sensors 23(16) : 7092, 2023
67. Sajjad M, Ullah FUM, Ullah M, Christodoulou G, Cheikh FA, Hijji M, Muhammad K, Rodrigues JJ : A comprehensive survey on deep facial expression recognition : challenges, applications, and future guidelines. Alexandria Engineering Journal 68 : 817-40, 2023
68. Ullah FUM, Obaidat MS, Muhammad K, Ullah A, Baik SW, Cuzzolin F, Rodrigues J, Albuquerque VHC : An intelligent system for complex violence pattern analysis and detection. Int J of Intelligent Sys 37(12) : 10400-22, 2022
69. Tutuianu GI, Liu Y, Alamäki A, Kauttonen J : Benchmarking deep Facial Expression Recognition : An extensive protocol with balanced dataset in the wild. Engineering Applications of Artificial Intelligence 136 : 108983, 2024
70. Mejia-Escobar C, Cazorla M, Martinez-Martin E : Towards a Better Performance in Facial Expression Recognition : A Data-Centric Approach. Computational intelligence and neuroscience : 1394882, 2023
71. Manierka MS, Rezaei R, Palacios S, Haigh SM, Hutsler JJ : In the mood to be social : Affective state influences facial emotion recognition in healthy adults. Emotion 21(7) : 1576-81, 2021
72. Shahzad T, Iqbal K, Khan MA, Imran, Iqbal N : Role of Zoning in Facial Expression Using Deep Learning. IEEE Access 11 : 16493-508, 2023
73. Tanioka T, Smith MC, Osaka K, Zhao Y : Framing the Development of Humanoid Healthcare Robots in Caring Science. Int J Hum Caring 2) : 112-120
74. Pepito JA, Ito H, Betriana F, Tanioka T, Locsin RC : Intelligent humanoid robots expressing artificial humanlike empathy in nursing situations. Nurs Philos 21(4) : 2020
75. Betriana F, Osaka K, Matsumoto K, Tanioka T, Locsin RC : Relating Mori's Uncanny Valley in generating conversations with artificial affective communication and natural language processing. Nurs Philos 22(2) : 2021
76. Tanioka T : The Development of the Transactive Relationship Theory of Nursing (TRETON) : A Nursing Engagement Model for Persons and Humanoid Nursing Robots. Int J Nurs Clin Pract 4 : 223, 2017