Lee J, Bennett CC, Stanojevic C, et al. (2023) Detecting cultural identity via robotic sensor data to understand differences during human-robot interaction. *Advanced Robotics.* 37(22), 1446-1459. DOI: 10.1080/01691864.2023.2277159

Detecting Cultural Identity via Robotic Sensor Data to Understand Differences during Human-Robot Interaction

Jinjae Lee[a]*, Casey C. Bennett[a,b]*, Cedomir Stanojevic[c], Seongcheol Kim[a], Zachary Henkel[d], Kenna Baugus[d], Jennifer A. Piatt[e], Cindy Bethel[d], Selma Sabanovic[f]

[a]*Department of Intelligence Computing, Hanyang University, Seoul, Korea;*
[b]*Department of Computing & Digital Media, DePaul University, Chicago, USA;*
[c]*Department of Parks, Recreation & Tourism Management, Clemson University, USA;*
[d]*Department of Computer Science and Engineering, Mississippi State University Starkville, USA;*
[e]*School of Public Health, Indiana University, Bloomington, USA;*
[f]*School of Informatics, Computing, and Engineering, Indiana University, Bloomington, USA;*

*Corresponding authors. Emails: jjlee93@hanyang.ac.kr; cabennet.iu@gmail.com;

Notes on contributors

Jinjae Lee is currently a Master's student in Data Science at the Department of Intelligence Computing at Hanyang University. He is interested in the application of machine learning and human-robot interaction to healthcare problems.

Dr. Casey Bennett is an Associate Professor in the Department of Intelligence Computing at Hanyang University in Seoul, Korea. He specializes in artificial intelligence and robotics in healthcare, including the use of data science and machine learning to create better human-robot interaction. He completed his Ph.D. at Indiana University in the US.

Dr. Cedomir Stanojevic is an Assistant Professor in the Department of Parks, Recreation & Tourism Management at Clemson University's College of Behavioral, Social and Health Sciences, SC, U.S.A. He specializes in recreational therapy and interventions related to leisure and improved quality of life, focusing on socially assistive robotics and ecological momentary assessment to improve various populations' health outcomes. He completed his Ph.D. at Indiana University in the US.

Seongcheol Kim received his Master's degree in Data Science at the Department of Intelligence Computing at Hanyang University. He is interested in the deep learning, natural language processing, and human-robot interaction to healthcare problems.

Zachary Henkel is a PhD student in the Department of Computer Science and Engineering at Mississippi State University. He is interested in human-robot interaction and related engineering related problems in social robotics.

Kenna Baugus is a graduate student in the Department of Computer Science and Engineering at Mississippi State University. She is interested in human-robot interaction and related engineering related problems in social robotics.

Dr. Jennifer Piatt is an Associate Professor in the Department of Health and Wellness Design at Indiana University- Bloomington, School of Public Health. She specializes in recreational therapy and interventions related to community-based rehabilitation. Focusing on Socially Assistive Robotics as a therapeutic intervention, sge aims to understand how emerging technologies can address clinical outcomes.

Dr. Cindy Bethel is the Billie J. Ball Endowed Professor of Engineering in the Department of Computer Science and Engineering at Mississippi State University. She is the director of the Social, Therapeutic, and Robotic Systems (STaRS) Lab, focused human-robot interaction and therapeutic robotic pets.

Dr. Selma Šabanović is a Professor of Informatics and Cognitive Science at Indiana University Bloomington, where she directs the R-House Human-Robot Interaction Lab. Her work combines the social studies of computing with research on human-robot interaction and social robotics. She explores the design, use, and consequences of socially interactive and assistive robots in diverse social and cultural contexts, including various countries, homes, schools, and healthcare environments. She completed her PhD in Science and Technology Studies at Rensselaer Polytechnic Institute.

Detecting Cultural Identity via Robotic Sensor Data to Understand Differences during Human-Robot Interaction

Socially-assistive robots (SARs) have significant potential to help manage chronic diseases (e.g. dementia, depression, diabetes) in spaces where people live, averse to clinic-based care. However, the challenge is designing SARs so that they perform appropriate interactions with people who have different characteristics, such as age, gender, and cultural identity. Those characteristics impact how human behaviors are performed as well as user expectations of robot responses. Although cross-cultural studies with robots have been conducted to understand differing population characteristics, they have mainly focused on statistical comparisons of groups. In this study, we utilize deep learning (DL) and machine learning (ML) models to evaluate whether cultural differences show up in robotic sensor data during human-robot interaction (HRI). To do so, a SAR was distributed to user's homes for three weeks in the US and Korea (25 participants), while collecting data on the human activity and the surrounding environment through on-board sensor devices. DL models based on that data were able to predict the user's cultural identity with roughly 95% accuracy. Such findings have potential implications for the design and development of culturally-adaptive SARs to provide services across diverse cultural locales and multi-cultural environments where users' cultural background cannot be assumed *a priori*.

Keywords: human-robot interaction; deep learning; cross-cultural robotics; adaptive robot design; human activity recognition

Word Count: 5875 (not including tables, figures, references)

1. Introduction

1.1. Background

In existing cross-cultural studies in the field of human-robot interaction (HRI), some research has suggested that there is no significant difference across cultures in the way humans and robots interact, while other studies conversely have shown the exact opposite [16-19]. Generally speaking, many of those studies focused on statistical comparisons to identify differences between cultures. However, the question exists whether statistical comparisons are the right way to evaluate those differences? Alternatively, we might instead ask whether proposed cultural differences in HRI would be detectable in robotic sensor data collected on second-by-second basis during human interactions over time. If there are significant differences across cultures, then one might hypothesize that they would manifest in how the same interactive behavior is performed by people from two different cultural backgrounds, even if only subtly, and thus should show up differently in the robotic sensor data. If so, we should theoretically be able to do the reverse of statistically analyzing differences based on known user cultural identity. **Rather, we should be able to develop models that can *predict* the user's cultural identity based purely on the robotic sensor data, even if the identity of a particular user is unknown.** Indeed, in multi-cultural settings (something increasingly common in the modern world), the robot may not *a priori* know the cultural identity of a particular user. Our goal here is to test the above hypothesis to evaluate its feasibility. To do so, we focus specifically on robots designed for healthcare purposes in home environments.

Socially Assistive Robots (SARs) are platforms designed to provide companionship and/or therapeutic benefits to patients outside clinical settings like hospitals, similar in some ways to the benefits of human caregivers [1]. In recent years, advances in medical technology, such as SARs, have enhanced our ability to manage long-term chronic diseases that are considered difficult to treat with minimal side effects, resulting in prolonged human lifespan and a higher quality of life. However, there are limitations to the current technology-based treatment approaches for such diseases that typically require intensive ongoing care and attention over the long-term. For instance, mental health disorders and neurological diseases, such as depression and dementia, often have repeated cycles of improvement and exacerbation, wherein a wide range of irregular symptoms and resulting behaviors may appear sporadically depending on the patient's condition at any given moment in time. Such irregularity in patient status demands that SAR technology is flexible and adaptive to individual users. At the same time, hospitalization or frequent visits to the hospital to manage a disease in which symptom trajectory is unclear can be a great burden in terms of time and money for the patient and their families [2-4]. As such, more adaptive SAR technology is a necessity for reducing patient burdens. Recent epidemics such as COVID-19 have only reinforced the need for better home treatment models and remote treatment, particularly for older adults and patients with limited mobility that make frequent visits to medical facilities a challenge [5-7].

As one solution to the above issues, many types of SARs are currently being investigated experimentally in order to create better user adaptation in real-world settings. To do so, SARs are typically deployed in in-home environments, interact with users in a continuous fashion, and monitor their daily life through data collected via various sensors, with the ultimate goal to help care for patients as well as to support the analysis of their health condition status. From a clinical perspective, SARs have shown benefits in improving social skills in children with autism spectrum disorders (ASD) [8, 9] and managing mental diseases such as dementia or depression problems in older adults [10-13], among other applications.

One core component of this adaptive SAR approach lies in social interactions between the robot and the human, which provides users mental and social support to reinforce/sustain other ongoing treatment modalities (similar to a human caregiver or therapy pet) [14]. As such, SARs need to be able to reliably identify people's activities within the home and infer context-specific user intent of those behaviors in order to facilitate successful interaction. In previous studies, SARs with various sensors have been deployed to collect different types of potentially useful data, with machine learning (ML) and deep learning (DL) techniques applied to the collected robotic sensor data in order to identify human activities [15]. However, creating appropriate SAR responses in real-world settings depending on the contextual situation remains a challenge, particularly given the amorphous nature of the real world. For example, accurately identifying the behavior of a patient suffering from depression and then determining which response is most appropriate in the current situation to assist them is a challenge even for human caregivers. It is also unclear whether robots should perform different actions according to a user’s individual characteristics, such as their cultural background or ethnicity, or whether more universal responses would be appropriate.

1.2. Research goal

The goal of this study is to explore further evidence for the idea that the way people interact with robots is based on cultural setting and/or location. To do this, ML/DL techniques were used to train data collected by various sensors onboard a SAR deployed in user homes in different cultures, then test whether the models could distinguish the cultural identity of users performing a set of common in-home interaction behaviors with SARs. Detecting cultural identity holds potential to be used as a new means of creating adaptive behaviors for robots within the HRI field and, at the same time, can be expected to provide potential insights about the necessity to design robots in the future with different cultural-specific interaction behaviors.

This study is divided into 5 chapters including the current chapter, Introduction. Chapters 2 and 3 introduce Related works and Methods, and Chapters 4 and 5 deal with Experiments, Results, and Discussion.

2. Related works

2.1. SAR in the HRI research field

HRI is a research field related to understanding, designing, and evaluating robot systems for use by humans. There is a diverse range of domains covered by the field, including search and rescue, assistive and educational robots, entertainment, military and police, and space exploration [20]. Unlike the robots of the past, which were often focused on manufacturing/production tasks geared toward replacing human physical labor by performing simple and repetitive tasks, today robots are often designed to support humans via a broader array of social functions, such as by communicating with humans to assist in accomplishing collaborative tasks or providing companionship.

Along those lines, one primary area of use for such assistive robots is in the medical field [21]. For instance, Robinson et al. (2014) provide a broad overview of robots with the functions required by older adults during activities of daily living (ADL), such as walking assistance robots, home care robots, injury prevention robots, and rehabilitation exercise robots, or robots being studied [22]. A similar application is the use of SARs for managing mental health conditions and monitoring health status in user homes via human-robot social interaction, rather than physically assisting with problems related to physical health [21]. Previous research has explored using SAR companion for a wide range of services designed for various target audiences (older adults, children, patients, relatives, caregivers) and various health-related functions (in-home event notifications back to clinicians, video connection with caregivers, detection of dangerous situations like falls, cognitive stimulation, etc.) [23].

2.2. Human activity recognition and EMA

Similar to activity recognition with robots, there are many studies using smartphones [24, 25], wearable devices [26, 27], and IoT devices equipped with various sensors for recognizing human activities, e.g. using accelerometer sensor data to classify specific types of ambulation such as walking, standing, and going up and down stairs [24, 26]. Lara et al. provide a good overview of the broad range of activity categories that have been the recent focus of research on human activity recognition, such as movement, modes of transportation, phone use, and activities of daily living like eating or sleeping [28].

One open question related to the above is how we can best make human activity discernible to the SAR based on its sensor data. One possibility is to develop ML algorithms that can classify sensor data patterns into identifiable activity. However, to do so requires ground truth labels that clearly connect human behavior to the sensor data patterns, allowing for the training of ML classifier algorithms following standard approaches in the field [29]. In previous studies, recall-based data collection techniques have been used to generate such ground-truth labels, e.g. questionnaires, phone calls, or interviews collected at the end of each day, week, or even the end of the entire experiment [11, 30]. However, a fundamental problem raised in existing human activity recognition research is that recall-based data collection techniques rely heavily on the user's memory after the experiment, which can be biased by human perceptions (or even influenced by later positive/negative events after the activity in question) and thus are not always accurate [29, 31-34]. Of course, if a SAR only has to classify a few, simple activities, it may be less likely that recall bias will be a major issue, but in practice SARs often need to classify complex and diverse human activities in real-world settings. Given that, recall-based techniques that rely on human memory to collect data have the potential to cause major errors in the long run.

As an alternative to recall-based data collection, ecological momentary assessment (EMA) combined with SARs is currently being utilized in research studies in the HRI field and beyond [15, 27, 29, 31]. Unlike controlled environments that are artificially set up for experimentation, such as laboratories or hospitals, EMA-based data collection allows users to autonomously evaluate interactions in real-time, without direct observation by the researcher or doctor. Moreover, such interaction data can be obtained in real-world environments such as a home or workplace, serving as the basis for ground-truth labels of interactions in later analysis [15, 27]. The EMA method, which also can enable tracking of daily changes in health status even beyond the clinic walls, is a useful tool to relieve the constraints of healthcare-related research on SARs in terms of place, time, and cost [35].

2.3. Sliding window technique for catching human activities

One of the most widely used sensors for human activity recognition are accelerometers. Accelerometer sensors have the advantage of compact size and low cost and are thus embedded in numerous consumer electronic personal devices, including smartphones and wearable devices [36]. Furthermore, because movement patterns can be calculated through x/y/z coordinate values, accelerometers can be seen as essential for identifying many human activities where the amount/type of movement are indicative of that particular activity. To accomplish that, it is important to detect the patterns of how the x/y/z coordinates change over time, i.e. the sequence information within the motion data. For example, such patterns can manifest as changes of direction occurring in sequence or the frequency of motion.

The sliding window approach is a common method used to model temporal data, like the motion data described above for activity recognition. It is useful for detecting periodic activities such as walking or running or static activities such as standing and sitting. [37]. In general, sliding time windows are a fixed-size set of data (e.g. 5 seconds) that are used for modelling purposes, starting from the first data point and then "sliding" along the dataset at some interval. Typically the time windows overlap, so each data point appears in multiple windows [15, 26, 28]. A wide range of window sizes such as 1sec, 2sec, and 10sec have been explored for human activity recognition in previous studies [37], but there is still no definitive answer for the optimal window size for a given activity. We return to this particular topic in the Methods and Results of our research below.

2.4. Cross-cultural HRI Research

We would be remiss not to mention at least briefly the broader topic of cross-cultural research in HRI, for which there is much existing literature. In general, cross-cultural studies are intended to understand the differences that may occur between groups based on effects of the cultural environment on human behavior, as well as the interplay of culture with other demographic factors such as race, age, and gender. In the field of HRI, research has been conducted on whether the interaction between humans and robots leads to the same results across groups with different cultural characteristics [16-19, 38-41]. For instance, Andreasson et al. (2018) studied the difference in the way emotions are conveyed to the robot through the tactile sense between men and women, and there was no difference in the frequency or intensity of touch, but it was found that women interacted longer than men in certain emotions [16]. Elsewhere, Anzalone et al.

(2019) performed a comparison between children with Autism Spectrum Disorder (ASD) and the Typical Development (TD) group on behaviors that draw attention from robots and the results showed that the ASD group showed a significantly lower interest in the robot's behavior [18]. Another study investigated how robot therapy with autistic individuals differs depending on the cultural background between patients from Asian and Eastern European cultural backgrounds and found significant differences [19]. A particularly interesting example can be seen in a study that investigated differences in greeting practices by country and its relevance to HRI [42]. Compared to robots greeting with general gestures that are widely used around the world, greeting with each culture's native gesture was shown to make users feel closer to the robots and increase ratings of likeability towards the robot.

In summary, understanding cultural differences and their impact on HRI can potentially contribute to the improvement of robot design and the development of more personalized, culturally-adapted robot services in the future. Our study here contributes to this broader body of cross-cultural HRI research, though our focus is specifically on applications to SARs.

3. Methods

3.1. Experimental Protocol & Data Description

The data used in this study were based on an EMA sampling approach for social robot interactions described in Bennett et al. (2021) [29]. The dataset consisted of approximately 250 million data samples collected from 25 participants (Korea - 13, US - 12). However, technical issues with data collection leading to partial data loss caused us to exclude 4 people from this analysis, leaving us with 21 total participants analysed here. The participants were drawn from the general population aged 20-35 years old and roughly 70% female. In previous experiments though, we have detected no gender differences with these kinds of robots [15], while the age range was a limitation due to the study being run during the coronavirus pandemic (normally we work with older adults). For the experiment, a robot sensor collar (Figure 1) developed at Mississippi State University was worn on Hasbro's Joy-For-All robot companion cat (Figure 1) and deployed to the participant's home. The study followed a *convergent parallel mixed methods* strategy [43], where the focus is on collecting a large amount of detailed data about each user rather than sparse data amount many users. The experiment was conducted for 3 weeks, and during the experiment, sensor data was collected through the collar roughly 9 times per second, and real-time data on the interactions was collected through the ExpiWell EMA mobile app (https://www.expiwell.com) using an EMA sampling approach described below. We have made a cleaned-up version of the dataset publicly to the Dryad Repository for other researchers to use, which can be accessed at this link: https://doi.org/10.5061/dryad.tb2rbp078

Figure 1. Sensor Collar

Figure 2. SAR with sensor collar

The collar has built-in sensors that collect data such as movement (x, y, z coordinates), light, sound, and environmental conditions (temperature, air pressure, humidity, air quality). That sensor data was used as the input features for modelling in the study here, while the target class to be predicted was the cultural identity of each participant (Korean or US). Previous studies have shown that such sensor data can be used to detect specific human interaction behaviors with roughly 80% accuracy [15, 29]. However, the primary question here was whether that same sensor data could also predict the cultural identity of the user. **In other words, could the robot detect the culture of the user, based purely on its sensor data patterns.**

The description of key features used as input is shown in Table 1. In order to collect data on naturalistic interactions in real-time, a notification ping was sent to the users smartphone approximately 5-7 times randomly throughout the day (following a standardized EMA protocol [29]). The six interaction behavior *modalities* collected through EMA in this study were “Playing”, “Talking”, “Petting”, “Listening to Media”

(TV, radio, YouTube, etc.), and “Moving a robot”, based on previous research in this domain [29]. A total of 585 direct and indirect interactions with the robot were analyzed here, comprising roughly 146 hours of total data.

Table 1. Feature List

Category	Features	Description
Accelerometer	accX, accY, accZ	Motion amount from the accelerometer in x, y (lateral), and z (up/down) directions
Rotation	arc	Amount of rotational motion during interaction
Light sensor	light	Raw values from the light sensor
Sound sensor	audioLevel	Raw values from the sound sensor
Sound category	Quiet, Moderate, Loud	Specific sound categories detected based on sound sensor thresholds
Air quality sensor	IAQ, co2Equivalent, gasResistance, breathVocEquivalent	Raw values for indoor air quality
Environmental sensor	temp, pressure, humidity	Raw values for indoor environmental conditions
Orientation category	Portrait Up Back, Portrait Up Front, Portrait Down Back, Portrait Down Front, Landscape Left Back, Landscape Left Front, Landscape Right Back, Landscape Right Front	Specific orientation categories detected based on accelerometer thresholds

3.2. Exploratory Data Analysis

The data distribution for US and Korean participants for each interaction modality are shown in Table 2. The distribution of interaction types with the SAR is similar in both groups. However, there were a couple exceptions. In the US group, the proportion of moving robot pets was nearly three times that of Korea, and in the Korean group, the proportion of watching media (e.g. TV, radio, YouTube) was about twice that of the US group. Those differences likely relate to differences in home living environments and lifestyles between the US and Korea, which have been reported in previous studies based on OECD data [15, 44]. In short, lifestyle differences between the two locations lead to differences in the physical built environment, which subsequently effect how certain behaviors are performed and their frequency as well as indoor environmental conditions.

Table 2. Basic Information

Modality	US				Korea			
	interaction	**time (min)**	**sample**	**rate**	**interaction**	**time (min)**	**sample**	**rate**
Playing	9	135	98744	4.1%	15	225	105534	4.3%
Talking	35	525	397502	15.9%	39	585	274629	11.2%
Petting	77	1155	893629	35.0%	116	1740	816775	33.3%
Media	35	525	405797	15.9%	118	1770	830851	33.9%
Eating / Cooking	16	240	185188	7.3%	33	495	232382	9.5%
Moving it	48	720	562655	21.8%	27	405	190275	7.8%
Total	220	3300	2543515	100%	348	5220	2450446	100%

As an initial sanity check, we visualized the sensor data to compare the two cultural groups to see if there were noticeable differences visible to the human eye. Those results were mixed. There were differences between the two cultures in terms of accelerometer, humidity, air pressure, and gas resistance sensor data, but there were no differences between the two cultures in terms of light, sound, temperature, CO2, and IAQ sensor data. As an example, Figure 3 shows a three-dimensional visualization of the distribution of the accelerometer data (x, y, z values) for movement by several interaction behavior modalities. There seems to be a difference in the movement between cultures based on these visualizations, comparing the US (orange) and Korean (blue) participants. Regardless, a statistical comparison of group averages was not the point of this study, as we were focused on ML/DL modelling which we describe below.

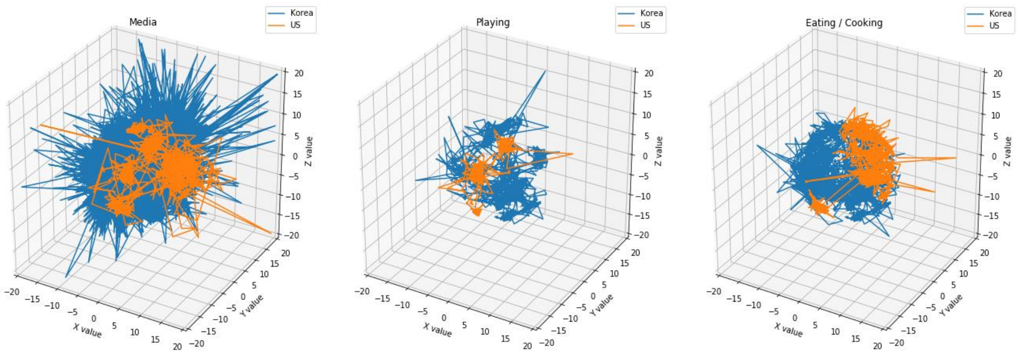

Figure 3. 3D visualization of data distribution (x, y, z values)

3.3. Preprocessing

The goal of preprocessing in ML/DL is to optimize the dataset in order to improve performance of the resulting models. Various preprocessing methods can be applied depending on the data used or how the problem to be solved is defined, such as removing error values, filling in missing values, dealing with outliers, and data scaling techniques (e.g. normalization). Several preprocessing methods were used in this study, which we detail in the subsequent sections.

3.3.1. One-hot encoding

Since a computer can only understand numeric values, it is necessary to convert data composed of other types into numeric data. Most of the features used in this study (Table 1) are numerical data, but since some fields such as orientation category and sound category were categorical data, a one-hot encoding process was performed to convert the categorical data to numerical data as shown in the example below (Figure 4).

Sound category
Quiet
Moderate
Loud
Moderate
Quiet
Quiet

Loud	Moderate	Quiet
0	0	1
0	1	0
1	0	0
0	1	0
0	0	1
0	0	1

Figure 4. One-hot encoding

3.3.2. Scaling

When using data as input to a machine learning or deep learning model, the model generally uses only the size for features without considering the units of each data. A large difference in the scale between different features or outliers that significantly deviate from the general data distribution can cause problems in the performance of ML/DL models. For example, if we were modelling home sale data, then the price of the home may be on the order of 0 to 1 million, while the number of bedrooms may be more like 1-4. Both are numbers, but the scales are vastly different, which can lead to some models over-emphasizing the home price feature at the expense of the number of bedrooms feature. Likewise, our dataset here included sensor features with vastly different scales, so normalization was performed to adjust the data range.

3.3.3. SMOTE

SMOTE is a widely used technique for resolving data imbalance problems. If the number of each target class used for model training is significantly different, most of the data will typically be classified as the most common class, which adversely affects the performance of ML/DL models. The problem can be mitigated through various class rebalancing methods or weighting schemes. In order to prevent class imbalance problems here, the input training data for modelling was adjusted using hybrid rebalancing methods known as SMOTE in order to produce a balanced dataset [45].

3.3.4. Dataset Restructuring via Sliding time windows

As described in Section 2.3, sliding time windows are commonly used to deal with temporal modelling problems. Human activities typically unfold over time (e.g. a few seconds or minutes) rather than occurring instantaneously. Think of sitting in a chair, for example. There is a sequence of movements that occurs over several seconds in order to go from standing up to sitting down. In order to make use of time series data with such time dependences in a predictive model, it is necessary to reconstruct the information in the form of a fixed window that can provide the model with “snapshots” of the data at a given point in time for modelling sequences of movement. However, in order to set an appropriate window size, one must consider how different activity types unfold. For example, during a 15-minute mealtime, a similar type of movement of picking up food and putting it in its mouth is repeated as one eats. If it takes about 3 seconds to pick up food and put it in the mouth, then it is appropriate to set the window size to 3 seconds to detect each instance of that activity. However, for a different activity that takes longer to unfold, such as petting an animal, a window size of 5 seconds or 10 seconds or even longer may be more appropriate. It is still an open question what window size would be best for detecting various human interaction activities during HRI.

As such, in this study, the amount and frequency of sensor data (i.e. “features”) collected at each time were considered for use of 8 window sizes (1, 2, 3, 5, 10, 20, 30, 60 seconds) for comparison purposes. The resulting data for analysis utilizing the sliding time windows can be visualized as a 3D matrix as shown in Figure 5, with each row of data forming the y-axis, the features forming columns in the x-axis, and the sliding windows being a third z-axis. Each interaction was essentially then represented as a cube and used as input feature data for modelling (Figure 6).

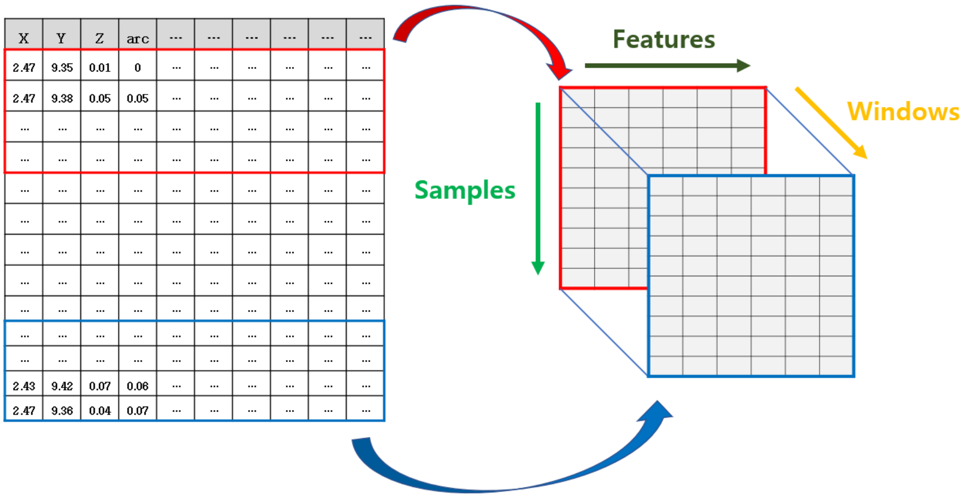

Figure 5. Creation of sliding time window

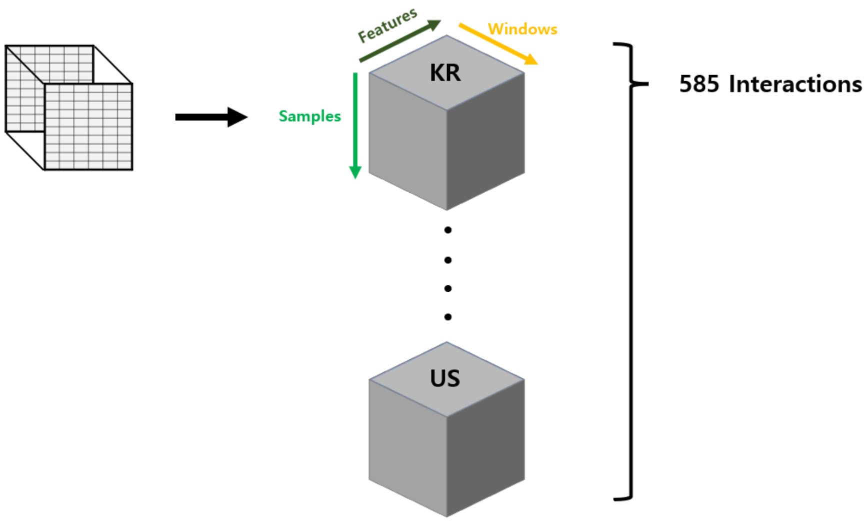

Figure 6. Example of input data for modeling (based on method in Figure 5)

3.4. ML/DL Modelling Approach

The modelling method used to predict the cultural identity divided each 15-minute-long interaction into a sliding window as in the prior studies performed in 2021 [29] and 2022 [15], with each window was set to overlap the data in the previous window by

50% [24, 25, 29]. An example of how the input “feature” dataset (for both ML and DL) were created from the raw sensor data is shown in Figures 5 and 6 in the previous section, though there was some subsequent post-processing for the ML analysis described below. To test whether such sensor data can be used to predict the cultural identity of the participant, the culture of the participant was coded as 0 (US) or 1 (Korea), which was then set as the “target” class for modelling. The target class was then appended on to the feature dataset, along with the interaction modality type of each interaction so that we could also later analyze specific interaction modalities separately for comparison.

ML models were built using the Python package SciKit learn, including random forests, gradient boosting, ada boosting, and support vector machines (SVMs). Models were generally run using the default parameters in Scikit. Results were evaluated using 5-fold cross-validation based on accuracy and AUC metrics, following standard ML guidelines [46]. For the ML models, the feature data cubes (described in Figures 5 and 6 in the previous section) were collapsed down into a single row of data per interaction, by taking averages or percentages/frequencies for each feature across all the sliding time windows across the 15-minute interaction period. Feature selection was also performed as part of the ML analysis, based on a wrapper method using random forest, which removes features of low importance relative to the target variable by building multiple models with different subsets of the features, then identifying a subset with the best performance.

DL models were constructed using the Python package Keras. DL models explored included recurrent neural networks (RNN) such as long-short tern memory (LSTM) and gated recurrent units (GRU), as well convolutional neural networks (CNN). To evaluate performance, 20% of the data was held out as a test set for each classification run, while the remain 80% of the data was used to train the models. The DL models used the feature data cubes described in Section 3.3.4 directly, rather than collapsing the data as was necessary for the ML models above. In terms of the DL model architectures more specifically, they comprised a series of layers (either RNN or CNN) up to 10 layers, with pooling layers in between as necessary. However, we found beyond just a few layers provided little performance benefit, so the DL results reported here are from those more parsimonious models. We also evaluated stacking RNN and CNN layers as combined models, in different combinations, as well as varying the unit size of layers and filter/kernel size. After experimentation, the optimal unit size for those RNN layers was determined to be around 200, while the optimal CNN layers were found to have filter size of 26 with kernel size of 8. In short, the final models generally comprised 5 layers, including an initial input layer, 2 RNN layers, 1 CNN layer, and a final Dense output layer using Softmax to make the final prediction.

We note that the order of the CNN and RNN layers varied for comparison, which are shown in the results below. Using DL model combinations with CNN layers first followed by RNN layers second did provide optimal performance, so those settings were subsequently used to evaluate the effects of different sliding time window sizes (ranging from 1 second up to 60 seconds) on compare model performance, which we describe in the 2nd half of the results below.

4. Results

While it is possible to perform a statistical comparison of the two cultures to evaluate differences during HRI, our analysis goal here was to test the feasibility of creating a predictive model that could detect cultural identity purely based on robotic sensor data, which could be of particular use for autonomous robots who are not yet sure with whom they are interacting. This section describes the prediction results of ML and DL models for cultural identity, including a comparison of different size sliding time windows and feature selection.

4.1. Cultural identity prediction

4.1.1. ML models

Table 3 shows the results of the various ML models. The model performance was fairly consistent across the different modelling methods, producing maximal accuracy scores around 90% with AUC of 0.94. Since collapsed data was used for the ML models, the reliability of the results was relatively low compared with the DL models, as indicated by the higher standard deviations of the accuracy and AUC scores (compare with table 5 in the next section).

Table 3. Cultural identity detection based on ML

Model	Accuracy	AUC
Random Forest	91% (+/- 8)	0.94 (+/- 0.07)
Gradient Boosting	90% (+/- 8)	0.94 (+/- 0.06)
Ada Boosting	91% (+/- 8)	0.94 (+/- 0.07)
SVM	86% (+/- 10)	0.87 (+/- 0.10)

Beyond that, the feature selection analysis found that 9 out of 43 features play an important role in predicting cultural identity based on robotic sensor data during HRI. Interestingly, the majority of those selected features (6 out of 9) were motion-related features (see Table 4). That may suggest that motion is more indicative of a user's cultural identity rather than sensor data related to light, sound, or other environmental conditions. In other words, there appear to be some detectable cultural differences in how users move when performing the same type of interaction modality. Some of those movement differences may relate to differences in physical home living spaces between Korea and the US mentioned in Section 3.2. For example, different physical layout might create different orientations of the person and robot during common interactions if there is less floor space, so that those interactions occur on furniture (table, bed, etc.) rather than at floor level. Or if the person is required to physically move the robot from room to room in a larger house. However, that may not fully explain movement differences within the same interaction modality, and there may something more subtle in terms of normal interpersonal distance or appropriate physical contact that varies across cultures affecting the human and robotic pet interactions here. More work would be needed (e.g. intensive case studies with a small number of users repeating the same behaviors over and over) in order to try to identify the specifics of those differences.

Table 4. Feature selection results (selected features)

	Feature
1	accX
2	accY
3	accZ
4	light
5	humidity
6	Moderate
7	Portrait Down Back
8	Landscape Right Back
9	Landscape Left Back

4.1.2. DL models

We also evaluated whether various DL models could accurately detect the cultural identity of human users during HRI. Results can be seen in Table 5. Those models generally produced accuracies and AUC scores in the mid-90s. The top-performing models combined CNN and RNN layers (CNN-LSTM, CNN-GRU), producing accuracy scores ~95% with AUC near 0.96. It has been previously reported with such HRI robotic sensor data that combining CNN and RNN layers in a single model provides performance advantages for theoretical reasons [29].

Table 5. Cultural identity detection based on DL

Model	Accuracy	AUC
LSTM	92.73% (+/- 2.66)	0.9581 (+/- 0.0268)
GRU	93.76% (+/- 2.23)	0.9756 (+/- 0.0097)
CNN-LSTM	94.87% (+/- 1.53)	0.9580 (+/- 0.0126)
CNN-GRU	94.96% (+/- 2.18)	0.9594 (+/- 0.0210)

Additionally, we were curious if for the top-performing DL models (in this case CNN-LSTM), whether it would be possible to predict the cultural identity based on individual interaction modalities by themselves, rather than the whole dataset. The results of that can be seen in Table 6. Those individual modality predictions averaged an accuracy of about 81% with AUC of 0.84, which was much lower than the overall performance of 95% accuracy reported in Table 5 above. The performances of individual modalities also ranged widely, from just 71% accuracy for the Talking modality to near 86% accuracy for the TV/Media modality. Regardless, though performance is suboptimal for the individual modalities relative to the overall predictive performance, 81% is not that bad for a binary prediction problem, and there may be situations where trying to identify a user's cultural identity during HRI based on a single interaction modality could be of utility.

Table 6. Cultural identity classification performance for each modality

Modality	Accuracy	AUC
Playing	82.00% (+/- 14.00)	0.8760 (+/- 0.1124)
Talking	70.67% (+/- 6.80)	0.7382 (+/- 0.0799)
Petting	84.64% (+/- 3.63)	0.8921 (+/- 0.0265)
Media	85.81% (+/- 3.59)	0.8811 (+/- 0.0382)
Eating / Cooking	77.00% (+/- 9.00)	0.7890 (+/- 0.0969)

Moving it	84.00% (+/- 10.41)	0.8756 (+/- 0.1030)
Average	**80.98% (+/- 7.91)**	**0.8420 (+/- 0.0761)**

Finally, we also compared the performance of different sliding time window sizes, to see if they impacted DL model performance (again based on CNN-LSTM). Eight window sizes were tested. A 5-second time window was used as a "baseline" in the models reported above, based on previous research [11,15,29]. So, the question here was whether other window sizes would perform better than the baseline. Results can be seen in Table 7. We note that the 1-second time window showed the lowest performance, while the 10-second time window showed the highest performance (Table 7). However, the 10-second window showed only marginal improvement beyond the 5-second window (95.4% vs 94.9% accuracy), which was not significantly different. We also note that for window sizes of 20 seconds or longer, performance gradually reduced. Given that, we can ascertain that there appears to be an optimal window size around 5-10 seconds for detecting cultural identity via robotic sensor data during HRI, at least for the set of interaction behaviors examined here.

Table 7. Exploring the optimal window size for performance improvement (CNN-LSTM)

Window size	Accuracy	AUC
1 second	90.86% (+/- 2.81)	0.9389 (+/- 0.0151)
2 second	91.37% (+/- 3.39)	0.9442 (+/- 0.0185)
3 second	92.65% (+/- 1.96)	0.9425 (+/- 0.0193)
5 second	94.87% (+/- 1.53)	0.9580 (+/- 0.0126)
10 second	95.39% (+/- 1.49)	0.9632 (+/- 0.0144)
20 second	94.19% (+/- 1.61)	0.9566 (+/- 0.0129)
30 second	94.36% (+/- 2.57)	0.9560 (+/- 0.0193)
60 second	93.93% (+/- 2.18)	0.9635 (+/- 0.0204)

5. Discussion Conclusion

5.1. Summary of Results

The purpose of this study was to evaluate whether it is possible to use ML/DL models to detect cultural identity of human users during HRI based purely on robotic sensor. Results showed that it was indeed possible using a dataset of 21 US and Korean participants and over 145 hours of in-home interaction data collected via EMA sampling techniques. A variety of ML and DL modelling techniques were evaluated, as well as different window sizes for temporal modelling. **The best performing models built on that dataset attained roughly 95% accuracy.** Interestingly, we also found that robotic sensors related to motion tended to be the most important in distinguishing between cultural identity, which suggests there may be something subtle about the way people move when performing the same interactive behavior modality that distinguishes their cultural background.

The idea behind this study originated from results of deep learning models for the two cultures in our own previous research [15], as well as questions arising from other researchers' cross-cultural HRI studies in the related work section (Section 2). In those studies, models generated to fit data collected *within* a specific culture suffered significant performance degradation when applied to other different cultures. In other words, the idea we can create universal behavioral models for in-home robotic pets or other social robots that will work anywhere in the world is not likely to be a successful strategy [47]. Generally speaking, we are all cognizant that there may likely be differences that occur during HRI depending on the cultural background of users. However, except for simple cases, such as eating behaviors where we see the use of chopsticks in East Asian countries versus the use of forks in Europe and US, there hasn't been much research on how cultural differences in the way a particular behavior (e.g. eating) manifests may impact interactive autonomous robots in users home and work spaces. Moreover, various cultural differences can also cause different ways of thinking resulting in different behaviors entirely.

One clear takeaway from the findings presented in this study is that not only is more research is needed, but the *right kind* of research. Currently, many cross-cultural HRI studies utilize statistical techniques to analyse differences in user behaviors across cultures at the population level, but there are limitations to that approach. Namely, it fails to evaluate the inter-individual differences in how the *same* behavior is performed by people of different cultural backgrounds, and only focuses on whether different groups perform behaviors with different frequencies. Therefore, in order to generate further evidence that the way people interact with robots is influenced on their cultural region or environment necessitates research like the kind performed here, where cultural identity predictions were attempted by *inversely* using data generated from interactions with robots to predict a user's cultural background. In particular, that allowed us to evaluate if that same interaction behavior performed by users from two different cultural backgrounds was significantly distinguishable, based purely on robotic sensor data. As the results showed here, they are indeed quite distinguishable, which potentially opens up new research opportunities to explore ways that autonomous robots can be designed to take advantage of that phenomenon.

5.2. Implications

This study has a number of implications for HRI researchers, across a variety of domains. For instance, in a healthcare setting, we could think about the implications for developing an adaptive SAR that is deployed for medical purposes for older adults who need continuous in-home care in daily life, such as depression or dementia. In particular, due to the nature of the medical field where even a slight error is difficult to tolerate in checking a patient's health condition, the ability to more precisely and accurately classify human behaviors based on their cultural background may be of use for remote monitoring or tracking fluctuations in the patient's activity levels and health status.

Keeping with that same example, currently medical data is often collected and utilized only within a specific region and/or country, but one could envision the utility to understand and utilize underlying cultural differences to provide global medical services beyond the scope of a specific country. Beyond robots, one could imagine using information drawn from behavioral data and sensor data from other types of devices in a similar manner to that described in this paper in order to detect a person's cultural identity, or otherwise to build an ML pipeline to classify input data by geographic locations or create a healthcare model optimized for a specific geographic location.

Within the domain of HRI more broadly, creating robots that can understand the cultural background of users holds promise for designing better interactive robots in the future. Such robots may be better able to autonomously adapt their behaviors, better aware of what the human is actually doing in particular scenarios, and thus ultimately perform more appropriate responses to humans over the long term [48]. To that end, one significant implication of this research is that doing so will require more research into the socio-cultural aspects that influence human behavior towards robots. Not only in a qualitative sense, but also related to how those socio-cultural aspects *subtly* affect data that roboticists may use to generate ML/DL models for autonomous robot behavior. **That may be of particular relevance in multi-cultural settings, where the cultural identity of the user may not be known to the robot *a priori*.** Undoubtedly, such multi-cultural environments are increasing in frequency in our modern globallyinterconnected world, and misunderstanding subtle social cues in such environments results in robots that are essentially socially inept [49]. In our view, that is currently a serious oversight amongst the HRI community.

5.3. Limitations

There were a number of limitations of this study which are important to keep in mind. First, one major limitation of the study was data-related. Originally the study was designed as a long-term in-home user study over the course of several weeks with each participant, featuring a total of 25 participants (12 US, 13 Korea). The plan was to collect all that data over a 6-month period. However, long-term in-home studies are notoriously challenging, and ours was no different. Due to a number of technical issues that caused partial data loss, as well as one participant's non-compliance with study protocol, we were left with 21 participants for this particular analysis.

Beyond that, a second challenge was the EMA sampling approach used. While EMA does produce more naturalistic data, it also results in more sparse data since one is only randomly sampling 5-7 times per day rather than continuously across time. That resulted in some sparsity in the analysis dataset used for the predictive models, where some activities had a limited number of examples despite data being collected for each participant over several weeks.

Third, since most of the participants in Korea and the United States were university students (mostly in their 20s), there are limitations in terms of the study population. Different age groups or socioeconomic groups may produce different results or interact with the robot in different ways. Therefore, it is necessary to collect data from more diverse populations, as that is known to impact HRI [15,16, 37, 39, 40]. It would also be interesting to compare immigrant groups living in other cultural environments (e.g. Asians living in the US, or vice versa), which is something we did not do in this particular study.

Finally, there are also some limitations in terms of experimental design. Various additional sensors could be used to collect surrounding environment data *off-board* the robot to provide a broader understanding user home environment. Similarly, sensor collar capabilities could be enhanced through interoperability with surrounding internet-of-things (IoT) devices, to extend their scope. Likewise, to accurately detect the user's movements, rather than using sensors only on the robot's collar, wearable devices could be deployed. We in fact ourselves are exploring integration of different technologies in ongoing research studies. In short, there is a myriad of possibilities. However, as other HRI studies have shown, we would note that generally does NOT include cameras in people's homes, as users find it uncomfortable and a serious privacy concern to have cameras or audio recording devices in their living space [50]. In regard to personal information protection, it is necessary to establish ethical standards to ensure the best interests of users while preventing data from being used for questionable purposes or shared indiscreetly with others [51, 52].

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by separate funding mechanisms in South Korea and the United States:

KOR - Hanyang University Research Fund, HY-2020

USA – National Science Foundation grant, IIS-1900683

References

[1] Liang A, Piroth I, Robinson H, et al. A pilot randomized trial of a companion robot for people with dementia living in the community. J Am Med Dir Assoc. 2017;18(10):871-878.

[2] Ospina MB, Krebs Seida J, Clark B, et al. Behavioural and developmental interventions for autism spectrum disorder: a clinical systematic review. PLoS One. 2008;3(11):e3755.

[3] Kang IO, Lee SY, Kim SY, et al. Economic cost of dementia patients according to the limitation of the activities of daily living in Korea. Int J Geriatr Psychiatry. 2007;22(7):675-681.

[4] Cantarero-Prieto D, Leon PL, Blazquez-Fernandez C, et al. The economic cost of dementia: a systematic review. Dementia. 2020;19(8):2637-2657.

[5] Lyth J, Lind L, Persson HL, et al. Can a telemonitoring system lead to decreased hospitalization in elderly patients?. J Telemed Telecare. 2021;27(1):46-53.

[6] Win AZ. Telehealth can bridge the gap for rural, disabled, and elderly patients. J Am Med Dir Assoc. 2015;16(4):268-269.

[7] Caratozzolo S, Zucchelli A, Turla M, et al. The impact of COVID-19 on health status of home-dwelling elderly patients with dementia in East Lombardy, Italy: results from COVIDEM network. Aging Clin Exp Res. 2020;32:2133-2140.

[8] Jain S, Thiagarajan B, Shi Z, et al. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. Sci Robot. 2020;5(39):eaaz3791.

[9] Scassellati B, Boccanfuso L, Huang CM, et al. Improving social skills in children with ASD using a long-term, in-home social robot. Sci Robot. 2018;3(21):eaat7544.

[10] Gross HM, Mueller S, Schroeter C, et al. Robot companion for domestic health assistance: Implementation, test and case study under everyday conditions in private apartments. In: Proceedings 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2015; Hamburg, Germany. p. 5992-5999.

[11] Bennett CC, Sabanovic S, Piatt JA, et al. A robot a day keeps the blues away. In: Proceedings 2017 IEEE international conference on healthcare informatics (ICHI); 2017; Park City, UT, USA. p. 536-540.

[12] Torta E, Werner F, Johnson DO, et al. Evaluation of a small socially-assistive humanoid robot in intelligent homes for the care of the elderly. J Intell Robot Syst. 2014;76:57-71.

[13] Šabanović S, Bennett CC, Chang WL, et al. PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. In: Proceedings of IEEE 13th international conference on rehabilitation robotics (ICORR); 2013; Seattle, WA, USA. p. 1-6.

[14] Rabbitt SM, Kazdin AE, Scassellati B. Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. Clin Psychol Rev. 2015;35:35-46.

[15] Bennett CC, Stanojevic C, Kim S, et al. Comparison of in-home robotic companion pet use in South Korea and the United States: a case study. In: Proceedings 2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob); 2022; Seoul, Korea, Republic of. p. 01-07.

[16] Andreasson R, Alenljung B, Billing E, et al. Affective touch in human–robot interaction: conveying emotion to the Nao robot. Int J Soc Robot. 2018;10:473-491.

[17] Libin AV, Libin EV. Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach. Proc IEEE. 2004;92(11):1789-1803.

[18] Anzalone SM, Xavier J, Boucenna S, et al. Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment. Pattern Recognit Lett. 2019;118:42-50.

[19] Rudovic O, Lee J, Mascarell-Maricic L, et al. Measuring engagement in robot-assisted autism therapy: A cross-cultural study. Front Robot AI. 2017;4:36.

[20] Goodrich MA, Schultz AC. Human-Robot Interaction: A Survey. Found Trends Commun Inf Theory. 2007;1:203-275.

[21] Feil-Seifer D, Mataric MJ. Defining socially assistive robotics. In: Proceedings of 9th international conference on rehabilitation robotics ICORR); 2005; Chicago, IL, USA. p. 465–468.

[22] Robinson H, MacDonald B, Broadbent E. The role of healthcare robots for older people at home: A review. Int J Soc Robot. 2014;6:575-591.

[23] Gross HM, Schroeter C, Mueller S, et al. Progress in developing a socially assistive mobile home robot companion for the elderly with mild cognitive impairment. In: Proceedings 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2011; San Francisco, CA, USA. p. 2430-2437.

[24] Ronao CA, Cho SB. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst Appl. 2016;59:235-244.

[25] Ogbuabor G, La R. Human activity recognition for healthcare using smartphones. In: Proceedings of the 2018 10th international conference on machine learning and computing (ICMLC); 2018; New York, NY, USA. p. 41-46.

[26] Casale P, Pujol O, Radeva P. Human activity recognition from accelerometer data using a wearable device. In: Proceedings of the 8th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA); 2011; Las Palmas de Gran Canaria, Spain. p. 289-296.

[27] Kim HJ, Lee SH, Lee SE, et al. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR mHealth and uHealth. 2019;7(10):e14149

[28] Lara OD, Labrador MA. A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutor. 2012;15(3):1192-1209.

[29] Bennett CC, Stanojević Č, Šabanović S, et al. When no one is watching: Ecological momentary assessment to understand situated social robot use in healthcare. In: Proceedings of the 9th International Conference on Human-Agent Interaction (HAI); 2021; Virtual Event, Japan. p. 245-251.

[30] Randall N, Bennett CC, Šabanović S, et al. More than just friends: in-home use and design recommendations for sensing socially assistive robots (SARs) by older adults with depression. Paladyn. 2019;10(1):237-255.

[31] Stanojevic C, Bennett CC, Sabanovic S, et al. Utilization of EMA and the Transtheoretical Model of Behavior Change to Prevent Relapse during Long-Term Socially Assistive Robot Based Interventions. In: Proceedings of LSI Workshop of the ACM/IEEE International Conference on Human-Robot Interaction (LSI-HRI); 2022; Sapporo, Japan.

[32] Duffy W, Curran K, Kelly D, et al. Addressing the problem of activity recognition with experience sampling and weak learning. In: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys); 2018; London, UK. p. 1238-1250.

[33] Appelboom G, Yang AH, Christophe BR, et al. The promise of wearable activity sensors to define patient recovery. J Clin Neurosci. 2014;21(7):1089-1093.

[34] Fridriksdottir E, Bonomi AG. (2020). Accelerometer-based human activity recognition for patient monitoring using a deep neural network. Sensors, 20(22), 6424.

[35] Bennett CC, Ross MK, Baek E, et al. Predicting clinically relevant changes in bipolar disorder outside the clinic walls based on pervasive technology interactions via smartphone typing dynamics. Pervasive Mob Comput. 2022;83:101598.

[36] Ni Q, Patterson T, Cleland I, et al. Dynamic detection of window starting positions and its implementation within an activity recognition framework. J Biomed Inform. 2016;62:171-180.

[37] Banos O, Galvez JM, Damas M, et al. Window size impact in human activity recognition. Sensors. 2014;14(4):6474-6499.

[38] Shahid S, Krahmer E, Swerts M. Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend?. Comput Hum Behav. 2014;40:86-100.

[39] Nomura T, Kanda T, Suzuki T. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. AI Soc. 2006;20:138-150.

[40] Rea F, Muratore P, Sciutti A. 13-year-olds approach human-robot interaction like adults. In: Proceedings of 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob); 2016; Cergy-Pontoise, France. p. 138-143.

[41] Morillo-Mendez L, Schrooten MG, Loutfi A, et al. Age-related Differences in the Perception of Eye-gaze from a Social Robot. In: Proceedings of 13th International Conference on Social Robotics (ICSR); 2021; Singapore, Singapore. p. 350-361.

[42] Shidujaman M, Mi H. Which country are you from? A cross-cultural study on greeting interaction design for social robots. In: Proceedings of 10th International Conference on Cross-Cultural Design. Methods, Tools, and Users (CCD); 2018; Las Vegas, NV, USA. p. 362-374.

[43] Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed methods Approaches, Thousand Oaks, CA, USA: Sage Publications; 2017.

[44] Oh JY, Bennett CC. The Answer lies in User Experience: Qualitative Comparison of US and South Korean Perceptions of In-home Robotic Pet Interactions. Proceedings of the 18th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI); 2023; Stockholm, Sweden. p. 306-310.

[45] Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321-357.

[46] Siebert J, Joeckel L, Heidrich J, et al. Towards guidelines for assessing qualities of machine learning systems. In: Proceedings of 13th International Conference on Quality of Information and Communications Technology (QUATIC); 2020; Faro, Portugal. p. 17-31.

[47] Bennett CC, Sabanovic S, Stanojevic C, et al. Enabling robotic pets to autonomously adapt their own behaviors to enhance therapeutic effects: A data-driven approach. In: Proceedings 32rd IEEE International Conference

on Robot and Human Interactive Communication (RoMAN); 2023; Busan, Korea, Republic of. In Press.

[48] Sabanovic S, Bennett CC, Lee HR. (2014). Towards culturally robust robots: A critical social perspective on robotics and culture. In: Proceedings of 9th ACM/IEEE International Conference on Human-Robot Interaction Workshop on Culture-Aware Robotics (CARs); 2014; Bielefeld, Germany.

[49] Chater N. How could we make a social robot? A virtual bargaining approach. Philosophical Transactions of the Royal Society A. 2023;381(2251): 20220040.

[50] Bartoletti I. AI in healthcare: Ethical and privacy challenges. In: Proceedings of 17th Conference on Artificial Intelligence in Medicine (AIME); 2019; Poznan, Poland. p. 7-10.

[51] Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics. 2021;22(1):1-5.

[52] Caine K, Šabanovic S, Carter M. The effect of monitoring by cameras and robots on the privacy enhancing behaviors of older adults. In: Proceedings of the 7th annual ACM/IEEE international conference on Human-Robot Interaction (HRI); 2012; Boston, MA, USA. p. 343-350.