An Intelligent Social Robot for Assisting with Multiple Daily Activities

by

Fraser Robinson

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Department of Mechanical and Industrial Engineering
University of Toronto

AN INTELLIGENT SOCIAL ROBOT FOR ASSISTING WITH MULTIPLE DAILY ACTIVITIES

Fraser Robinson

Master of Applied Science

Department of Mechanical and Industrial Engineering

University of Toronto

2023

Abstract

As the population ages, there is an increasing demand for support with activities of daily living (ADLs) in the homes of older adults and in long term care. Socially assistive robots (SARs) have shown success in increasing ADL independence by providing adaptive assistance for a variety of tasks including eating and dressing. The objective of this thesis is to develop new technologies for improving the design of SARs for ADL assistance. Namely, this thesis develops: 1) a novel social robot and wearable sensor system for assisting with the ADL of dressing, and 2) a new deep learning ADL recognition architecture for autonomously recognizing and monitoring known and unknown ADLs. Experiments evaluate performance using classification accuracy and real-time functionality using metrics such as usefulness and reliability. Results for classification performance show the developed methods outperform existing work while interaction experiments validate the systems for use with a variety of diverse users.

Acknowledgements

I would like to thank my supervisor Professor Goldie Nejat for her guidance and support throughout my degree. Thank you, Professor Nejat, for the many opportunities to learn new ideas, grow as a researcher, and experience robotics research in ways beyond what I could have asked for. Thank you to my ASBLab colleagues for engaging conversations and frequent support. Special thanks to Aaron Tan for welcoming me to the lab and to Zinan Cen, along with his supervisor Professor Hani Naguib, for being outstanding collaborators throughout my degree. Thank you to my committee members for your time and feedback. Thank you to my roommates, the community of St Jamestown, and Grace Toronto Church for comradery and council. Thank you to my partner Michelle for her patience and encouragement in all sorts of seasons. Thank you to my family; my brothers who taught me to have fun, my mom who taught me to value every type of person, and my dad who taught me to finish what I started. Finally, thank you to God, who knows deeply and puts all things into perspective.

Table of Contents

Table of Figures ... vii

Table of Tables ... viii

1. Introduction ... 1

1.1 Background ... 1

1.2 Socially Assistive Robots ... 2

1.3 Challenges ... 2

1.3.1 SAR Physical Features and Interaction Modes ... 3

1.3.2 The Need for SAR Adaptable Behaviors ... 3

1.3.3 Intelligent Autonomy ... 4

1.3.4 Challenges Summary ... 5

1.4 Thesis Objective ... 5

1.5 Summary of Contributions ... 6

1.5.1 Literature Review ... 6

1.5.2 A Social Robot and Wearable System for Dressing Assistance ... 7

1.5.3 ADL Recognition for SARs ... 7

1.5.4 Conclusions ... 7

2. Literature Review ... 8

2.1 SARs for ADL Assistance ... 8

2.1.1 Physical SAR Design Features ... 8

2.1.2 Older Adult Care Studies on SAR Features ... 19

2.1.3 SAR Awareness and Behavior Frameworks ... 21

2.1.4 Older Adult Care Studies on SAR Behaviors ... 23

2.1.5 Discussion on SAR Design ... 25

2.2 Assistive Technologies for Dressing ... 28

2.2.1 Non-Contact Assistive Technologies for Dressing 28

2.2.2 Assistive Technologies for Dressing using Robotics 30

2.3 ADL Recognition for SARs 31

2.3.1 Multimodal DL Networks for HAR 31

2.3.2 Embeddings of Feature Spaces for HAR 32

2.3.3 HAR for Social Robots 33

2.4 Chapter Summary 34

3. A Social Robot and Wearable System for Dressing Assistance 35

3.1 Robot Adaptive Behavior Deliberation 35

3.2 Performance Testing 39

3.2.1 Robot Adaptive Behavior Deliberation Cumulative Reward and Convergence .. 39

3.2.2 Dressing Step Classifier Accuracy 41

3.3 Demonstration Study 41

3.3.1 Participants 42

3.3.2 Procedure 42

3.3.3 Measures 43

3.3.4 Results 43

3.3.5 Discussions 45

3.3.6 Considerations and Limitations 46

3.4 Chapter Summary 47

4. ADL Recognition for SARs 48

4.1 Network Architecture 48

4.1.1 Video Backbone Network 49

4.1.2 Pose Backbone Network 50

4.1.3 Object Detection Backbone Network 51

4.1.4 Spatial Mid-Fusion Module 52
4.1.5 Dense Neural Layer 54
4.1.6 Transfer Learning 55
4.2 Architecture Training 55
4.3 Testing 56
4.3.1 Architecture Testing 56
4.3.2 Ablation Study 56
4.3.3 ADL Embedding Performance 57
4.4 Discussion 59
4.5 Chapter Summary 59
5. Conclusion and future recommendations 60
5.1 Summary of Contributions 60
5.1.1 A Social Robot and Wearable System for Dressing Assistance 60
5.1.2 Activity Recognition for SARs 60
5.2 Recommendations and Future Research 60
5.3 Concluding Statement 61
6. References 62

Table of Figures

Figure 1. Socially assistive robot-wearable system for dressing assistance proposed architecture [50]. 35
Figure 2. MAXQ socially assistive robot-wearable system dressing task graph [50]. 36
Figure 3. Cumulative reward of users for measuring task performance given by MAXQ and a hierarchical system without behavior preference learning [50]. 40
Figure 4. Dressing step classifier test accuracy in confusion matrix [50]. 41
Figure 5. Demonstration setup at technology conference [50]. 42
Figure 6. Box and whisker plot of questionnaire ratings for each question for all participants. Median is represented as bold lines IQR is represented by the boxes [50]. 44
Figure 7. Box and whisker plot of questionnaire ratings between females and males. Median is represented as bold lines IQR is represented by the boxes [50]. 45
Figure 8 - Proposed DL ADL recognition and classification architecture [51]. 48
Figure 9 - Video backbone network architecture [51]. 50
Figure 10 - Pose backbone network architecture [51]. 51
Figure 11 - Object detection backbone network architecture (YOLOv5) [51]. Adapted from [169]. 52
Figure 12 - Spatial mid-fusion, ADL embedding vector, and dense neural layers modules for ADL classification [51]. 53
Figure 13 - Pose spatial reshaping and scaling operations. 53
Figure 14 - ADL embedding spaces [51]. 58

Table of Tables

Table 1. Existing SARs for older adults [49]. 13
Table 2. Behavior utterance and gesture examples [50]. 36
Table 3. Behavior utterance and gesture examples [50]. 37
Table 4. Study questionnaire organized by attribute with descriptive statistics [50]. 43
Table 5 - Model modality test accuracy [51]. 57
Table 6 - Intra-class variation and inter-class distance for embedding spaces [51]. 58

CHAPTER 1

1. INTRODUCTION

1.1 Background

An increasing number of adults globally are requiring assistance to complete activities of daily living (ADLs), fundamental tasks required to care for oneself including personal hygiene, eating, and dressing [1]. Difficulties in completing ADLs can occur due to mild to moderate cognitive impairments [2] or functional limitations [3] which are known to increase with the overall aging process [3]. By 2050 the population of adults 60 years of age and older is expected to double to 2.1 billion, and those 80 and older to triple [4]. The resulting increase in older adults who require support to complete ADLs will cause a greater number of older adults needing to transition to living in long-term care (LTC) homes, which provide 24-hour onsite professional care to support their physical and cognitive needs [5]. Consequentially, the demand for an already dwindling healthcare system is expected to grow substantially [3] with an estimated caregiver shortage of more than 100,000 workers in the US alone by 2030 [6]. This lack of staffing combined with a new environment can lead to individuals feeling isolated from social circles, often worsening existing conditions such as dementia [7] and decreasing overall quality of life (QoL) [8]. There exists an urgent need for innovation in ADL assistance to improve quality of life and overall wellbeing, and to help address the strain on the labor force, and the various needs of a diverse aging population [9]. Solutions must be multifaceted, adaptive, and sustainable, and need to be supported by government policies and programs that also consider socioeconomic factors that affect health to meet both urgent and future older adult care needs.

A proposed care method to assist with ADLs while maintaining QoL and independence is reablement [10]. Reablement is a person-centered approach for gaining or regaining skills required to complete ADLs [10]. It consists of goal-directed rehabilitation interventions in which caregivers encourage and motivate individuals to expand their capabilities through coaching [11]. This strategy moves caregivers away from the "*do for*" norm (i.e., doing the ADL for the older adult) to the "*do with*" approach (supporting the older adult as they complete the ADL) [12]. Deployment of reablement programs have shown significant improvements in health and ADL ability in addition to decreased care costs for a variety of individuals including older adults [13], [14] and stroke patients [11]. However, the global shortage of caregivers has limited the implementation of these programs [15], and assistive technologies have not yet been incorporated. In addition to challenges associated with cognitive and physical decline of older adults, assistive solutions must be accessible for users who are novices in newer forms of technology [16].

1.2 Socially Assistive Robots

Socially assistive robots (SARs) are a unique type of robotic technology that use social communication modes to engage with people [17]. SARs have the potential to aid caregivers by assisting older adults in the completion of ADLs and assessing changes in ADL ability over time all in a single multi-facet technology [18]. The unique ability of SARs to adapt their behaviors to older adults can help support their individual needs and preferences as they age, making SARs especially well-suited for reablement care [19]. This adaptability combined with the potential efficiency of SARs to help advocate for appropriate caregivers-to-older-adults ratios in LTC homes aligns with policies aimed at meeting the health and social needs of older adults created by a shortage of caregivers in order to provide personalized quality care [20].

Existing work has used SARs to assist older adults with a small number of ADLs including dressing recommendations [21], exercise [22], or meal eating and cognitive games [23]. Crucial to having positive human-robot interaction (HRI) experiences in older adult care settings is the overall design of the robot, considering the unique challenges and opportunities that come with novice users in specific task and environment contexts [24]. SAR design includes aspects of physical appearance, interaction modes, and behaviors. Studies on SARs for assisting older adults with ADLs have found user preferences exist for each aspect of the design however there exist design features that remain either underexplored, underdeveloped, or both [25].

1.3 Challenges

There are several open challenges to designing SARs as effective long-term ADL assistants for older adults. These challenges include both design decision challenges (e.g., how to determine what the robot should do) and technology development challenges (e.g., enabling the robot to do something it cannot currently do).

1.3.1 SAR Physical Features and Interaction Modes

Individual user preferences need to be considered in the design of SARs, while maintaining feasibility for mass production and deployment. SARs can range in such appearance characteristics as human-likeness, size, expressiveness, and material composition. Interaction modes include social interactions via verbal and non-verbal communication, gestures, displays, and physical touch. A mixed-methods approach in [26] used a combination of questionnaires, interviews, and focus groups to determine the older adult expectations towards a hypothetical SAR for everyday assistance. Questions on preferred height, exterior finish, and favorite overall appearance (from a list) showed no one option was most preferred. Additionally, older adults expressed their desire to understand the functionality of the SAR before forming an opinion on its appearance.

Accommodating preferences is important in HRI as personalization increases engagement and enjoyment, having a positive impact on overall use by older adults [27]. Existing SAR designs for older adults have shown significant differences in appearance and interaction modes even when robots perform the same task [17], [25], [28] suggesting diverse expectations have challenged SAR developers to optimize robot design. In accommodating such preferences there is a risk in underrepresenting the diversity of users including of racial, cultural, gender, and age minorities, which can create inherent biases, i.e. similar to some medical voice dictation systems being more accurate for men than women [29]. The complexity of this challenge is increased by the changes in behavioral and attitudinal response when comparing those that directly engage in HRI with physically embodied robots to those who are asked their opinions on images or videos of SARs [30]. Researchers must determine which features are important based on user abilities and interaction context, while ensuring SAR accessibility to a broad and diverse userbase of older adults. Additionally, the development of new modalities (e.g., use of wearable sensors for motion tracking) remains an open technical challenge.

1.3.2 The Need for SAR Adaptable Behaviors

SAR behaviors can encompass varying strategies from emotional [31] to persuasive [32], while also considering social norms [33] to engage with older adults. An open challenge exists to adapt robot behavioral strategies to achieve the expectations of older adults and gain user trust and adherence [34]. Focus groups of older adults in [35] were presented with an imaginary scenario that put in conflict adherence to SAR recommendations to promote independence and older adult autonomy in disobeying the SAR suggestions. The study highlighted the expectations for SARs to have adaptive behaviors that consider user emotions and engagement as well as long-term user patterns in schedules and moods. Designing behaviors to consider social norms presents challenges in determining which cultures to consider. Caution must be taken to avoid ageist views of selected norms [36]. Beyond potential demographic biases, ethical concerns in developing SAR behaviors for older adults include: 1) privacy over recording user data to influence behavioral adaptation, 2) transparency of SAR intent, and 3) user autonomy in situations where a SAR attempts persuasion [37].

Development of adaptive behaviors requires training of learning methods to [38]: 1) detect and classify user state, 2) determine an appropriate SAR behavior, and 3) learn from user responses. Advancements in AI including machine and deep learning (DL) methods can improve the robustness of SAR behavior adaptation frameworks to changes in older adult behaviors overtime [39], which may occur due to cognitive decline [40]. Recently deployed SARs with adaptive behavior frameworks are yet to offer a holistic solution that both synthesizes a wide range of available data on the user state and applies this data to modify behavioral strategies accordingly [38]. The combination of inter-group and intra-user variability requires SAR behavior adaption that considers user preferences and cognitive changes while maintaining reliable task performance.

1.3.3 Intelligent Autonomy

Current deployments of SARs in senior care vary in their control architectures from teleoperation scenarios [27], where a human operator (visible or non-visible to the users) must be present, to full autonomy [25], where a robot is capable of HRI without expert human intervention. For long-term use, autonomy is the only sustainable option and to be achieved SAR architectures need to directly incorporate user(s), robot, and task environment information. For older adults, cognitive decline can decrease their ability to express thoughts using typical sentence structures [41] or facial expressions [42], limiting the use of standard natural language processing (NLP) and facial expression detection methods for interpreting user state. To account for the inexperience of older adults with robotics, SARs must provide alternate means of maintaining core functionality in the presence of hardware failures such as leveraging multimodal interaction modes using sensor fusion techniques [38].

Although historically robots in manufacturing only needed to be proficient in a single repeated task in a structured environment [43], autonomy for SARs in senior care is further complicated by the multiple tasks older adults expect them to reliably perform [44],[45] in diverse environments from kitchens to bedrooms in private homes [46] to common dining and recreational rooms in LTC [47]. SARs applications need to handle high environment variability and learn to adapt to their users' abilities and needs, while dealing with sensing uncertainty or unpredictable human behavior.

1.3.4 Challenges Summary

Designing SARs to assist individuals with ADLs includes design features of physical appearance, interaction modes, and behaviors. Physical appearance has been shown to be responsible for a variety of HRI outcomes, presenting a challenge in determining its relative importance to different users, activities, and scenarios [48]. Interaction modalities must be easy to use for individuals who are novice users of technology and several types of interaction modalities for social robots have yet to be developed. SAR behaviors should be adaptable to improve both the user intent to use and overall task performance. Intelligent autonomy for SARs must be robust to variations in users, different activities, and diverse environments to enable long-term use.

1.4 Thesis Objective

The objective of this thesis is to design, develop, and test autonomous social robots for ADL assistance. For SARs to successfully act as long-term helpers for individuals requiring ADL assistance, they must be easy to interact with, adapt their behaviors to suit user preferences, and be capable of autonomously observing and acting on user behaviors to initiate assistive HRI. To consider each of these challenges, the objective of this thesis is two-fold. Namely, this work focuses on the development of two unique SAR architectures: 1) a social robot wearable system for dressing assistance, developing a novel smart clothing interaction modality and assistance approach using adaptive SAR behaviors for the ADL of dressing, and 2) a new multimodal DL architecture for SARs to autonomously recognize and monitor multiple ADLs performed by different users in diverse environments and proactively initiate assistive HRI.

To enable a novel smart clothing for user motion monitoring, resistive strain sensors are developed using a unique manufacturing process. Processing of the sensor signals from the smart clothing is accomplished using a series of DL networks to map user motion to user joint angles and user joint angles to specific dressing steps (e.g., put right arm through, button up). A SAR behavior adaptation module is developed to provide social assistance to the user as they perform the dressing activity and modify SAR behavior using a MAXQ reinforcement learning (RL) architecture with the reward of user compliance and engagement. Experiments are conducted to assess the quantitative performance of each module in addition to the overall system and a demonstration-based user study evaluates the perceptions of key stakeholders towards the system.

The development of a multimodal DL ADL recognition network for SARs includes the selection of three backbone networks for extracting modality specific features from RGB-video, user 3D pose locations, and object locations from an RGB-image. A novel spatial mid-fusion method is developed to synthesize complimentary semantic features from each modality with a universal spatial understanding. A user motion embedding space isolates ADL motion from non-ADL motion to enable real-time deployment without the need to pre-define unexpected user motions not related to ADLs. A low dimensional representation of ADL features, defined as an ADL embedding space, is developed to contextualize similarity between ADLs observed offline during training and ADLs observed online during implementation. Comparison experiments evaluate the network's classification performance compared to unimodal and dual modal approaches. ADL embedding experiments evaluate the ability of the architecture to contextualize new ADLs not seen during training.

1.5 Summary of Contributions

This thesis seeks to contribute the development of novel technologies for designing SARs as ADL assistants. To this end, the contributions proposed are discussed in the following sections.

1.5.1 Literature Review

Chapter 2 first presents a review on social robots designed to help older adults to understand the current technologies and design strategies that exist for social robots within the context of older adult care. After recognizing key enablers and barriers to the deployment of social robots designed for use by older adults, this chapter then discusses literature specific to 1) the ADL of dressing, including assistive technology for dressing, smart clothing, and wearables used with SARs, to highlight the opportunities of using a SAR for dressing assistance, and 2) activity recognition, both as a general field and also developments specific to SARs.

1.5.2 A Social Robot and Wearable System for Dressing Assistance

Chapter 3 details the design, development, and testing of a social robot wearable system for dressing assistance. This chapter introduces a novel smart clothing wearable designed to facilitate interaction with a social robot based on the motions of the user. User motion is classified into an activity state defined as the most recently completed dressing step using DL networks and the robot responds to the user state with an adaptive behavior selection architecture that learns user preferences during the interaction. Testing is performed to evaluate 1) the performance of the system in correctly mapping user motion to joint angles, classifying joint angles to dressing steps, and adapting robot behavior based on the user response, and 2) the evaluation of potential stakeholders using a demonstration-based user study evaluating factors including perceived ease of use and usefulness.

1.5.3 ADL Recognition for SARs

Chapter 4 details the design, development, and testing of a multimodal DL architecture for SARs to recognize and assist with multiple known and unknown ADLs in real-time. This chapter includes details on each of the modality specific backbones, the development of the spatial mid-fusion method for obtaining the ADL embedding space, and the methods used to train the model. An ablation study is completed to show the classification accuracy improvements achieved using the developed multimodal approach compared to other unimodal and dual-modal methods. An evaluation of the ADL embedding space is performed and the results discussed for their ability to enable distinguishing between seen, unseen, and atypically performed ADLs.

1.5.4 Conclusions

Chapter 5 discusses the contributions of this work with respect to the development of SARs for ADL assistance using novel interaction modes and ADL recognition methods to enable improvements in SAR autonomy. Future recommendations are also presented to extend this work.

CHAPTER 2

2. LITERATURE REVIEW

This chapter will discuss the existing literature relevant to SARs for ADL assistance. First, from the work published in [49], an overview is presented of SARs developed for older adults, including those specifically for ADL assistance, in terms of their physical appearance, interaction modalities, and behavior selection. Next, using the review conducted in [50], the focus is placed on the ADL of dressing to discuss relevant assistive technologies to understand the unique opportunities that exist for using SARs. Then, from the review performed in [51], SAR intelligence is discussed in more detail including methods for ADL recognition and user state estimation relevant to enabling SAR autonomy.

2.1 SARs for ADL Assistance

This review is from work first published in [49].

2.1.1 Physical SAR Design Features

The main physical design features to consider when developing SARs for older adults are: 1) overall robot appearance, and 2) interaction modes. Each feature is discussed within the context of promoting effective social HRI and improving health and wellbeing outcomes while aging.

SAR Appearance

In general, older adults have specific, yet varying, preferences for the appearance of SARs which aid in increasing trust, perceived competence, and acceptance of these robots [17]. These attributes can be classified as human-likeness, expressiveness, size, and material composition.

Human-likeness: The appearance of a SAR may be classified as: 1) human-like, 2) character-like, 3) machine-like, or 4) animal-like, depending on the body and face features. *Human-like* robots have similar human facial features including eyes, eyebrows, a nose, and a mouth, and body features including a torso and two arms; *Character-like* robots have rounded heads and bodies, with minimal features, such as a face with only eyes. *Machine-like* consists of varying heads and body shapes ranging from square to rectangular with components including parts and linkages exposed; and *Animal-like* robots have shapes resembling those of the animals they mimic with many possessing fur.

An example of a human-like SAR from the waist-up is Brian 2.1 [23] which has a torso with a waist and two arms to promote familiarity and a silicone face with two eyes, eyebrows, a mouth, and a nose that can deform to display facial expressions. Brian has been used to assist older adults in LTC with cognitive interventions including memory games [23] and meal eating [52]. Milo R25 is a human-like SAR similar in appearance to a small child having an elastic frubber (foam + rubber) face with two eyes, eyebrows, a mouth, and a nose [53]. Milo R25 has been used to provide conversation therapy to older adults living with Alzheimer's disease [53]. Alice, an older version of Milo R25 was deployed in aging-in-place to support older adults with depression [54].

Character-like SARs with a combination of a head and arms include 1) Pepper [55], Casper [46], ARI [56], Stevie [57], Bandit [58], NAO [59], and Mini [60] which all have a rounded face with eyes and a mouth and a torso with two arms, and 2) Hobbit a one-armed robot with a head consisting of only eyes [61]. Character-like SARs with a head but without arms include Pearl [62] and iCat [63] (mouth and eyes), and Kompai [64] and Max [65] (eyes but no mouth). Some applications of character-like SARs are ADL assistance such as Casper for meal assistance [39]; cognitive stimulating games with Stevie [66]; monitoring for falls and providing calendar reminders using Max [67]; and exercise facilitation with NAO [59],[68], and Bandit [58].

Tangy is an example of a machine-like robot due to its square face and torso, and its visibly exposed cables. Tangy has been used to facilitate group-based cognitive interventions like Bingo [69] and Trivia [70]. Baxter [71], used for exercise, is also machine-like with its large frame, square head and exposed cables. Companion robots such as the popular seal-like PARO [72] and cat-like JoyForAll Cat [73] are animal-like as they resemble real life animals in shape and texture and are used for older adult pet-therapy to address loneliness and depression.

Expressiveness: Focusing on non-verbal visual expressiveness through embodiment, SARs may be classified using any combination of: 1) gaze direction, 2) facial expressions, 3) gestures, and 4) head and whole-body poses. Hobbit [61] uses head pan and tilt rotations to adjust its gaze direction. Max [67] displays both gaze direction and facial expressions through its LCD eyes by changing eye direction, color, and shape. iCat [63] actuates its head, mouth, eyes, and eyebrows for gaze direction and facial expressions. Pearl [62] is able to blink its eyes. PARO [72] and JoyForAll Cat [73] can blink and use head and whole-body movements to show emotions. Mini [60] and Stevie [57] have gaze direction, facial expressions using animated eyes, and head and whole-body movements. Other robots including Brian 2.1 [23], Milo R25 [53], Pepper [74], Casper [46], ARI [56], Bandit [58], NAO [59], Tangy [69], and Baxter [71] use all four types of visual expression for a variety of tasks such as exercise [58] and games [75] to promote user engagement [31].

Size: SARs can be classified by three different height ranges: 1) small-size (< 100 cm), 2) mid-size (100-125 cm), and 3) large-size (125-170 cm). Small-size SARs include desktop robots NAO [59], Mini [60], Milo R25 [53], and iCat [63] in addition to companion robots PARO [72] and JoyForAll Cat [73], the latter of which are similar in size to the animals they resemble. Pepper [55], Casper [46], Pearl [62], Kompai [64], Max [65], Bandit (when on a mobile base) [58], and Hobbit [61] are all mid-sized SARs and have been deployed in a wide variety of interactions. Large-size SARs are ARI [56], Brian 2.1 [23], Tangy [69], Baxter [71], and Stevie [57] which are near the average female height of 165 cm [56], are all deployed in LTC for cognitive interventions [56] ADL assistance [23], and games [57],[69].

Material Composition: The materials used to develop the robot's outer-shell/casing include: 1) hard plastic, 2) metal, or 3) soft materials. All types of shells are used to prevent robot damage from external factors. SARs with hard plastic shells are Pepper [55], Casper [46], ARI [56], Stevie [66], Pearl [62], Kompai [64], Max [65], Hobbit [61], Bandit [58], Baxter [71], iCat [63], and NAO [59]. Tangy [69] has an aluminum structure which suits its machine-like appearance. Soft materials include silicone for Brian 2.1 [23] and a custom formed frubber for Milo R25 [76] to emulate artificial skin. Fabrics including artificial fur on Mini's torso [77] and on the outer layers of PARO [78] and JoyForAll Cat [73] customize appearance and texture, and promote physical touch.

Interaction Modes

Interaction modes describe the interfaces SARs use to communicate with older adults including speech [55], sounds [79], visual displays [65], gestures [58], and physical touch [72].

Speech: Speech is important for SARs interacting with older adults as it provides them a familiar and intuitive form of bidirectional communication [80]. SARs may be classified based on their capability to: 1) speak, 2) detect spoken keywords, and 3) detect word associations (sentences). Some SARs can only speak such as iCat [63], Tangy [69], and Bandit [58]. Other SARs that speak also recognize certain keywords to initiate, pause, or end tasks such as Pearl [62], Kompai [64],[81], Max [65], Mini [60], and Hobbit [61]. SARs capable of both speech synthesis and recognition include Pepper and NAO with their built in NAOqi Natural Language Processing (NLP) [47], Brian 2.1 using Julius [23],[82], and Casper using IBM's Watson [39],[83]. ARI [56] and Stevie [57] have built in speech modules yet to be implemented in HRI studies with older adults. To-date, acoustic models and training data for NLP specific to older adults is limited [38] and standard available NLP software is less accurate due to differences in voice acoustics [84], cognitive function [41], and sentence structures [85] for this user group.

Sounds: SARs use sounds proactively or reactively to express robot states such as sleep, wakefulness, or excitement to increase engagement [78]. PARO [72] and JoyForAll Cat [73] make sounds such as cooing or meowing at various volumes and tones for pet-therapy [79]. Mini [77] uses sounds such as laughter, whistling, and yawning. Due to cognitive decline, non-verbal vocalizations like "hmm-mm" or "ugh" are more frequently used by older adults to express themselves [41], however these sounds have yet to be used as input for HRI.

Gestures: Human gesture types include [86]: 1) illustrators that add emotional expression and emphasis to speech (i.e., body language), 2) manipulators used subconsciously that involve interaction between body parts or other objects like fidgeting, and 3) emblems used deliberately to represent words like head-nods or head-shakes. SARs for older adults focus mainly on displaying and detecting illustrator and emblem gestures. SARs that use illustrator gestures include Brian 2.1 [23], Pepper [55], Casper [46], ARI [56], Bandit [58], NAO [59], Mini [60], Stevie [57], and Tangy [69] to indicate focus of attention when speaking [69],[70] or complement the emotion in speech [23]. Brian 2.1 [23] can determine user engagement by detecting illustrative gestures based on Canadian cultural norms. Pepper [47] uses emblem gestures such as bowing and waving to display cultural competency specific to either Japanese or British backgrounds. Hobbit [61] uses emblem gestures for different commands such as swiping for menu navigation. Bandit [58], Baxter [71], and NAO [59],[68] use emblems during exercise tasks to communicate proper exercise form and detect user compliance.

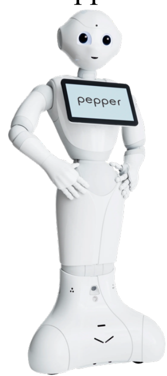
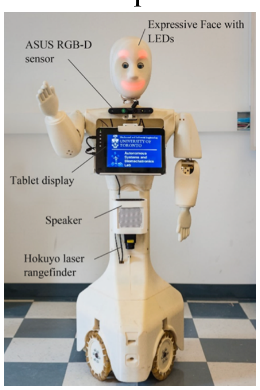
Displays: Visual displays are used to provide task specific instructions [46], show pictures or videos [47], or for teleconferencing with other people [67]. Displays may be output only or interactive touchscreens. Tangy uses its torso display to show Bingo numbers and Trivia questions to augment its speech [69],[70]. SARs with touchscreens include Casper to provide meal assistance instructions and offer recipe choices [46], Stevie for voice/video calling [57], and Pearl to add upcoming appointments to its calendar [62]. Furthermore, Pepper [47], Kompai [64], Max [65], Hobbit [61], Mini [60], and ARI [75] all display cognitive games on their touchscreens, which is especially valuable for older adults where comprehension speeds will vary [41]. Display height is typically targeted to accommodate older adults in a seated position, some displays may be tilted to improve accessibility when standing [55],[64],[65].

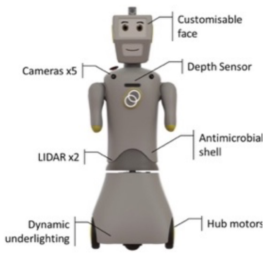
Physical touch: In general, SARs do not touch a user however some can detect physical touch. Physical touch detection may be categorized [87] as: 1) affective, for showing appreciation, 2) instrumental, to achieve a specific task, 3) controlling, to get attention, and 4) ritualistic, for greetings or departures such as handshakes. SARs that detect affective touch such as petting or stroking include the pet-like robots JoyForAll Cat [73] and PARO [88]. Baxter [71] detects instrumental touch during interactive exercise games. Mini [77] responds to controlling touch (e.g. a tap on head) for initiation of tasks. Pepper [55] detects ritualistic touch from sensors, i.e., on the top of its head, as a means of putting the robot in/ out of sleep mode. Culture was not explicitly considered in developing physical touch for SARs; however, it could help to promote generalizability to older adults with different cultural backgrounds.

Table 1 presents a summary of the design features and applications of the aforementioned SARs, with respect to the categories for type of appearance and interaction mode. In general, studies on the effectiveness and efficacy of SARs for older adults have shown positive outcomes most notably in cognitive training, ADL assistance, and as multifaceted solutions to prolong aging-in-place [17], [25], [89]. These applications are the most common for the SARs in Table 1. The use of SARs for social and psychological therapy requires more rigorous testing as current studies are limited to short-term interventions and their results are influenced by external factors such as changes in the daily lives of older adult [89].

Table 1. Existing SARs for older adults [49].

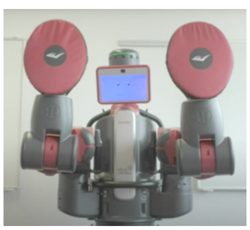
SAR	Applications	Appearance Categories	Interaction Modes	Works
Brian 2.1  Courtesy of ASB Lab, UofT	Engages in cognitive and memory games, meal-eating assistance	**Type**: Human-like **Face**: Gaze direction, facial expressions through deformable face **Body**: gestures, and head and upper - body movements **Size**: Height: 135 cm **Outer Shell Material**: Silicone face, aluminum body	**Input**: Sentence recognition, affect detection through body poses and gestures, wearable and object-based task specific sensors, illustrator gestures **Output**: Speech synthesis, facial expressions illustrator gestures	[28][23][52]
Milo 25  Courtesy of Robotkind	Conversation therapy for older adults with Alzheimer's disease	**Type:** Human-like, **Face:** Gaze direction, facial expressions through deformable face **Body:** gestures, and head and upper - body movements **Size:** Height: 50 cm **Outer Shell Material**: Polymer face, hard plastic body	**Input:** Sentence recognition **Output:** Speech synthesis, illustrator gestures	[53][54]

Pepper  Courtesy of RobotLAB	Engages in conversations, facilitates games and exercise	**Type:** Character-like **Face:** Different eye colors for showing emotion **Body:** gestures, head and whole-body poses **Size:** Height: 120 cm **Outer Shell Material**: Hard plastic (injection molded)	**Input:** Sentence recognition, touchscreen, ritualistic touch on head to sleep **Output:** Speech synthesis, illustrator and emblem gestures, touchscreen	[47][55] [74][90] [91]
Casper  Courtesy of ASB Lab, UofT	Assists with meal preparation	**Type:** Character-like **Face:** Gaze direction, facial expressions through LEDs **Body:** gestures, head and whole-body poses **Size:** Height: 125 cm **Outer Shell Material**: Hard plastic (3D Printed)	**Input:** Sentence recognition, touchscreen, **Output:** Speech synthesis, illustrator gestures, touchscreen	[39][46] [92]
ARI  ©PAL Robotics 2021, all rights reserved	Provides reminders for scheduled activities, cognitive games, fall detection, audio/video calling,	**Type:** Character-like **Face:** Gaze direction, different head colors for showing emotion **Body:** gestures, head and whole-body poses **Size:** Height: 165 cm **Outer Shell Material**: Hard plastic	**Input:** Sentence recognition, touchscreen, **Output:** Speech synthesis, illustrator gestures, touchscreen	[56][75]

Stevie  Courtesy of Trinity College Dublin	Engages in conversations, facilitates group games	**Type:** Character-like **Face:** Gaze direction, facial expressions through LCD display **Body:** head and whole-body poses **Size:** Height: 140 cm **Outer Shell Material**: Hard plastic	**Input:** Sentence recognition, touchscreen **Output:** Speech synthesis, illustrator gestures, touchscreen	[57][66]
Bandit  Courtesy of USC Interaction Lab	Physical exercise coach, cognitive games	**Type:** Character-like **Face:** Gaze direction, facial expressions through actuated eyebrows and mouth **Body:** gestures, head and whole-body poses **Size:** Height: 110 cm **Outer Shell Material**: Hard plastic	**Input:** Exercise emblem gestures **Output:** Speech synthesis, exercise emblem gestures	[58][93]
NAO  Courtesy of RobotLAB	Smart home interface, exercise coach	**Type:** Character-like **Face:** Gaze direction, different eye colors **Body:** gestures, head and whole-body poses **Size:** Height: 58 cm **Outer Shell Material**: Hard plastic	**Input:** Sentence recognition, exercise emblem gestures **Output:** Speech synthesis, illustrator gestures, exercise emblem gestures	[59][68][94]

Mini  Courtesy of UC3M Robotics Lab	Cognitive games, interactive dance	**Type:** Character-like **Face:** Gaze direction, facial expressions, and head and whole-body poses **Size:** Height: 50 cm **Outer Shell Material**: Hard plastic with fur clothing	**Input:** Sentence recognition, touchscreen, controlling touch for starting tasks **Output:** Speech synthesis, illustrator gestures, touchscreen	[60][77]
iCat  Courtesy of Christoph Bartneck	Engages in conversations, provides reminders and weather information	**Type:** Character-like **Face:** Gaze direction, facial expressions **Size:** Height: 38 cm **Outer Shell Material**: Hard plastic	**Output:** Speech synthesis	[63]
Hobbit  Courtesy of Hobbit Project	Provides reminders, household object retrieval, cognitive games, fall detection, exercise	**Type:** Character-like **Face:** Gaze direction **Size:** Height: 125 cm **Outer Shell Material**: Hard plastic	**Input:** Keyword recognition, command emblem gestures, touchscreen **Output:** Speech synthesis, touchscreen	[61][95]

Pearl  Courtesy of CMU Robotics	Provides reminders, mobile navigation guide	**Type:** Character-like **Face:** Blink **Size:** Height: 120 cm **Outer Shell Material**: Hard plastic	**Input:** Keyword recognition, touchscreen **Output:** Speech synthesis, touchscreen	[96][62]
Kompai  Courtesy of Kompai	Provide reminders, mobile navigation guide, audio/video calling, cognitive games	**Type:** Character-like **Face:** Static **Size:** Height: 125 cm **Outer Shell Material**: Hard plastic	**Input:** Keyword recognition, touchscreen **Output:** Speech synthesis, touchscreen	[64][81]
Max  Courtesy of Ilmenau University of Technology	Provide reminders, smart home interface, audio/video calling, cognitive games, fall detection	**Type:** Character-like **Face:** Gaze direction, eye-color to show current task, eye shape to show state **Size:** Height: 120 cm **Outer Shell Material**: Hard plastic	**Input:** Keyword recognition, touchscreen **Output:** Speech synthesis, touchscreen	[65][67]

Tangy  Courtesy of ASB Lab, UofT	Group cognitively stimulating activities	**Type:** Machine-like **Face:** Gaze direction, facial expressions through mouth actuation **Body:** gestures, head and whole-body poses **Size:** Height:140 cm **Outer Shell Material**: Aluminum	**Input:** Keyword recognition, task progress through RGB-D camera **Output:** Speech synthesis, illustrator gestures, display	[32][69] [70][97]
Baxter  Courtesy of CORiS, OSU	Exergames	**Type:** Machine-like **Face:** Gaze direction, facial expressions on LCD display **Body:** gestures, arm poses **Size:** Height: 178 cm **Outer Shell Material**: Hard plastic	**Input:** Exercise emblem gestures, instrumental touch **Output:** Exercise emblem gestures	[71]
PARO  Courtesy of PARO Robots	Pet therapy	**Type:** Animal-like **Face:** Blinking **Body:** Head and whole-body movements including head shaking, tail and flipper movements **Size:** Height: 16 cm **Outer Shell Material**: Hard plastic with fur cover	**Input:** Affective touch **Output:** Expressive sounds and movements	[72][78] [79][88] [98][99]

JoyForAll Cat Joy for All  Courtesy of JoyForAll	Pet therapy	**Type:** Animal-like **Face:** Blinking **Body:** Head and whole-body movements including head nodding and paw raising **Size:** Height: 26 cm **Outer Shell Material**: Hard plastic with fur cover	**Input:** Affective touch **Output:** Expressive sounds and movements	[73][99]

2.1.2 Older Adult Care Studies on SAR Features

Several HRI studies have been conducted with older adults to measure and compare appearance features and interaction modes of SARs using key concepts such as trust, likeability, and intent to use. In general, these studies either have low participation numbers (e.g., between 5-10 users) [67],[81],[92] or are limited to a single interaction [63],[71],[78]. This is primarily due to limitations in working with vulnerable populations such as older adults with dementia who may face cognitive fatigue when engaging in such research studies [47]. However, critical user trends within these studies can still be identified with respect to such measures as trust, intent to use, and enjoyment, and can be used to inform other similar studies. It is important to note that this field of HRI is still in its infancy [17], while also considering the challenges of working with vulnerable populations and the novelty of the SARs being tested [100].

Human-likeness: In [92], the influence of robot embodiment in assisting with a tea-making ADL on the overall perceptions and experience of HRI for older adults with mild cognitive impairments (MCI) was investigated. Three different platforms were used consisting of a character-like robot (Casper), machine-like robot (Ed), and a tablet placed on a table. Questionnaire results showed that Casper was the most preferred and engaging robot due to its dynamic features.

In [101], three robot characteristics were individually manipulated: robot face (none, machine-like, character-like), voice (none, digitized, human), and interaction mode (none, display tablet, touchscreen) to determine their influence on the affect of older adults during medication delivery. A character-like face and a touchscreen had the most influence on self-reported positive affect with the latter also increasing engagement as measured with heart rate. Using a human voice also increased positive affective response, however, with a lower effect size.

Expressiveness: In [63], the iCat character-like robot was used to explore the combined effect of facial expressions (smiling and nodding/none) and gaze (looking at user/not looking at user) on SAR acceptance by older adults during an information providing task (weather, reminders, etc.). Participants who interacted with the expressive iCat showed more conversational expressions, however, this did not increase SAR acceptance as measured by post-interaction surveys.

Size: In [77], the 50 cm tall SAR Mini was placed in a LTC home for 2 months where residents could freely interact with the SAR to engage in exercises using its touchscreen. Questionnaires for older adults, caregivers, and relatives showed high scores for usefulness, ease of use, and satisfaction with the SAR being perceived as friendly, smart, and safe. However, older adults did not believe Mini could increase their autonomy.

Material composition: In [78], PARO was used with independent living older adults to explore potential emotional support benefits. Older adults participated in a guided introduction to PARO and its capabilities with opportunities to hold and interact with the SAR using touch. Post-interaction interviews showed that older adults most liked PARO's fur, color, and cute appearance while they least liked PARO's limited functionality including inability to understand speech.

Verbal Communication: In [81], Kompai was deployed in the homes of older adults living alone. Users could ask the SAR to perform several tasks using either verbal communication or a touchscreen. Users below 80 years of age had no clear preference of communication mode, however, there was a significant preference for adults older than 80 to use speech.

Sounds: In [73], older adults engaged with the JoyForAll Cat in their own homes during two months to investigate if the robot could decrease loneliness. Older adults reported a decrease in loneliness, and interviews showed they appreciated the presence of the SAR. However, while the SAR could make sounds, many older adults noted the lack of interaction and responsiveness.

Gestures: SARs that can show and understand illustrative gestures have been rated highly by older adults for intent to use and enjoyment [23]. Expressing emblem gestures using Pepper to display cultural competency can increase the emotional wellbeing of older adults as shown in HRI studies with this robot using culturally appropriate greeting gestures such as bows and waves with users from both Japan and Britain [47].

Displays: The addition of touchscreens increases positive emotional response and engagement [101] and supports verbal information from SARs through the use of text and visuals in tasks like Trivia [70] or cognitive games [77]. Furthermore, a separate study with Hobbit showed speech and a touchscreen were significantly preferred to understanding emblem gestures for giving the SAR commands [54].

Physical Touch: In [71], the machine-like Baxter robot was used to determine the effect of physical touch (hitting/none) on older adult enjoyment of exercise games. Participants completed 8 games with varying amounts of physical touch in the form of hitting pads mounted to force sensing actuators. All activities that involved hitting rated high for enjoyment.

In [67], the SAR Max was deployed in residential care apartments to allow older adults to use features such as medical reminders, audio/video calling, and emergency detection. Some participants reacted emotionally to its behaviors, speaking to it as a social entity and frequently touching the SAR during interaction. Physical touch has been shown to also increase moods in pet-therapy sessions with PARO [78].

2.1.3 SAR Awareness and Behavior Frameworks

To design adaptive behaviors and intelligent autonomy for SARs, robot architectures have been developed to achieve robot awareness (task and user classification) and personalization (adaptive behaviors) based on the varying needs of older adults.

Task and User State Classification

Task classification is used to identify and monitor the steps needed in completing a particular task. User state classification considers user affect and engagement throughout the interaction as input for robot behaviors. Both classification forms may use data from onboard robot sensors including RGB-D cameras [69],[102] or user and object sensors [52],[103].

Task State Classification: In [102], task classification was performed by the character-like HomeMate SAR using RGB-D and laser scan data with a dynamic Bayesian network (DBN) to determine observed task states between the ADLs of meal preparation, cooking, eating, and taking medication. The SAR extracted features from the skeleton model of older adults and the relevant objects, i.e., dishes or a fridge for the DBN to classify the most likely task observed.

In [69], Tangy autonomously facilitated Bingo games with LTC residents. The game progression for a player was monitored when the older adults requested personalized assistance using an infrared reflective system detected by the Robot's IR sensor. Once the SAR approached the older adult, a 2D camera was used to detect the Bingo card features using its unique identifier picture and determine the location of number markers to classify the Bingo card as: 1) marked correctly, 2) incorrectly marked and/or missing markers, and/or 3) a winning card.

User State Classification: In [103], the Pepper robot classified user valence and arousal into an affect detection model using multilayer perception neural networks during a robot emotion elicitation activity. The affect of older adults from a LTC home was measured using an EEG sensor during robot emotional dancing which was defined as upper body movements to express either positive valence and high arousal or negative valence and low arousal based on movement speed and dynamics [103].

In [104], NAO used RGB cameras to identify facial features for emotion classification between seven different expressions using a Random Forest Classifier. Using a combination of distance, polygonal area, and elliptical area features resulted in good accuracy with older adults even when their faces were partially occluded by fingers or glasses.

Task and User State Classification: In [52], Brian 2.1 classified the progression of the meal eating task with older adults using a smart tray (with embedded force sensors) and utensil tracking system (using Wiimotes onboard the robot) to provide appropriate prompts, social encouragement and reinforcement. Body language and face orientation were also tracked and classified using a 3D Kinect sensor to determine if the older adult was distracted or accessible during meal eating to reengage them if needed using different robot emotions.

Adapting SAR Behaviors

Behaviors of autonomous SARs were initially designed with finite state machines (FSMs) to provide predefined responses for sets of identified inputs from users and their environments [27]. Recently, adaptive behavioral control has been used in SARs via: 1) robot self-learning through direct interactions with users [47], 2) learning from demonstrators (e.g., caregivers or experts)

[97], or 3) a combination of the two [39]. In designing SAR behaviors for effective HRI with older adults, emotional [31] or persuasive [32] strategies can be implemented to adapt to user and task specific preferences [33].

Task Behavior Learning Methods: Reinforcement learning (RL) methods have been used for SAR self-learning using rewards such as level of engagement [47]. However, since all behaviors must be attempted for a SAR to learn those that elicit high rewards, there is a risk that older adults may need to repeat a negative response to poorly received behaviors several times, causing confusion or frustration [39]. Learning from demonstration (LfD) has been used for SARs to learn new skills by directly observing them from caregivers, such as Tangy learning to autonomously facilitate Bingo sessions from caregivers in LTC homes [97]. LfD benefits from fewer user interactions to determine behavioral strategies, however, it can rely on a significant number of demonstrations. A unique hybrid approach using both LfD and RL can also be used to provide robot learning of task-specific behaviors through LfD and then personalization through online learning using RL, as was used by Casper in assisting older adults to make tea [39].

Emotional Behaviors: SARs use emotional models to communicate their intent and internal states to older adults during assistance to improve older adult understanding of the robots [31]. FSMs use transition rules to relate inputs from SAR sensors to robot emotion state changes to improve older adult task performance [23],[68]. Alternatively, in [91], an *nth* order Markov Chain based emotion model was developed for the Salt robot to determine when to display the four emotions of happy, interested, sad and worried, in response to user engagement in an activity, user affect and the robot's own emotional history.

Persuasive Behaviors: Persuasion in HRI seeks to change users' attitudes or behaviors [96]. Persuasion strategies used by SARs when interacting with older adults can be categorized as: 1) motivation strategies [93], and 2) compliance gaining persuasive strategies [32]. Persuasion strategies are frequently used by assistive technologies to achieve compliance and engagement in ADLs by older adults and present opportunities for similar strategies to be used by SARs [105].

2.1.4 Older Adult Care Studies on SAR Behaviors

Studies have been conducted with older adults to investigate various robot behavior learning methods and the use of SAR emotion and/or persuasion on HRI experience. User and state classification have been mainly used as inputs for SAR behavior adaptation.

Task Behavior Learning Methods: In [47], Pepper used RL to determine discussion topics and robot gestures based on user engagement, measured by older adult verbal responses, in order to improve user emotional states. The SAR's dialogue was personalized overtime as users engaged in conversations and playing games with the robot. Conversation personalization was focused on British and Japanese cultural topics. After two weeks of interactions, emotional wellbeing improved compared to a baseline group.

In [97], Tangy used LfD to learn from caregivers how to autonomously facilitate Bingo sessions using behaviors including calling out Bingo numbers and checking Bingo cards. Teachers could further customize the robot's learned behavior by modifying the SAR actions using a graphical user interface. An HRI study with LTC residents showed that older adults believed the SAR was easy to use and found Tangy's behaviors helpful and enjoyable.

In [39], Casper used LfD to learn assistive behaviors from allied-healthcare students from nursing, occupation and physical therapy, and speech-language pathology, to assist older adults in the tea making activity using verbal and non-verbal -based prompts with varying levels of speech directness (assertive/suggestive) and movement activity (high/medium/low). Casper's behavior was further personalized using on-line RL based on completion of activity steps. User studies with Casper [92] and residents in a retirement home showed they perceived Casper as socially intelligent and had high levels of engagement and positive affect.

Emotional Behaviors: In [23], Brian 2.1 played a matching card memory game with older adults in LTC setting. The robot displayed emotional behaviors (happy, neutral, sad) using an FSM-based behavior model that autonomously determined voice and facial expressions based on player accessibility (high to low). Questionnaire results showed that emotional expression was the most liked feature of Brian 2.1, which also received high scores for enjoyment and acceptance.

In [106], the Salt robot autonomously facilitated exercise sessions for older adults living in LTC. The robot guided the participants through multiple repetitions of upper-body exercises, using its *n*th order Markov model to determine its emotional response (happy, interested, sad, worried). The majority of users maintained a positive valence throughout the sessions with Salt and believed their physical health was improved. They were also motivated to continue performing daily exercises with the robot after the 2-month study was completed. Both emotional behavior adaptation studies were based in Canada [23], [106].

Persuasive Behaviors: In [93], Bandit was used to facilitate a musical cognitive game with older adults to explore whether it could improve cognitive attention through adaptive motivational behavior. Bandit changed its assistance level, using an FSM based on user reaction time and the percentage of game questions answered incorrectly, between 1) no hints, 2) directing when to press a button, and 3) saying which button to press. Analysis of older adult engagement during the activity confirmed the SAR was able to maintain user attention and improve task performance.

In [74], Tangy used a Thompson Sampling based approach during Bingo game facilitation to learn a personalized persuasive strategy for encouraging a specific older adult to comply with requests for playing. The persuasive strategies learned included neutral, praise, suggestion, and scarcity. A user study with Tangy and residents of a LTC home was conducted to explore engagement based on visual focus and compliance during group gameplay [69]. Tangy's personalized assistance was found to increase engagement with all users having very high compliance with SAR requests.

2.1.5 Discussion on SAR Design

Does the appearance of the robot matter? In comparison studies that focused on assistive tasks, SARs with more human-like appearance received higher ratings for engagement, perceived intelligence, and intent to use [92]. Studies using simplistic character-like SARs suggest that robot capabilities, namely the tasks performed [81] and interaction modes enabled by its appearance such as gestures [67], were the main design aspects that older adults were concerned with over appearance. This focus on capabilities over appearance has also been shown in focus groups on assistive robots with older adults [107].

Typically, preference studies have focused on showing pictures and videos of SARs [26],[108] instead of physical robot interactions, further limiting the real in-person experiences of older adults. Additionally, the considered works do not specify any cultural differences in appearance preferences and there are also no quantifiable differences between studies due to large intra-study variation. To fully understand appearance preferences requires long-term studies deploying and comparing SARs with similar capabilities but varying appearance types to isolate appearance effects while considering demographic and culturally diverse users. It is suggested that in developing SARs for older adults, functionality and familiarity are very important to this user group and should be the main priorities during feature design. In this thesis, development is done using the robot Leia (based on the Nao platform by Softbank robotics) for its gesture and speech functionality and character like appearance to provide familiarity.

How many interaction modes is too much? When available, verbal communication was found to be the most used and liked interaction mode with older adults for social interactions [109] and task commands [81]. The presence of sounds for pet therapy [73], gestures for exercise [58], touchscreens for cognitive games [77], and physical touch for exergames [71] were also found to positively influence HRI. Current works have not directly considered cultural differences in interaction mode preferences, however, some have customized modes such as greeting gestures based on culture which has shown to further improve HRI [47]. Cultural customization can be applied to both verbal and non-verbal interaction modes such as spoken expressions or using symbolic representations on interfaces. SARs with multiple interaction modes provide older adults accessibility and flexibility in using the modes that best suit their physical limitations and personal preferences which improves HRI [61]. An open challenge is to determine when the cost of adding additional interaction modes outweighs the benefit to the older adult users who may not use these modes [61] or find them annoying [68].

SAR developers should focus on identifying and improving existing highly valued interaction modes with this population, such as verbal communication, which numerous studies claim as being well-liked [52],[81] but also dysfunctional [27] due to existing audio and speech issues. There are also opportunities to explore less commonly used interaction modes in new contexts or develop new interaction modes to meet the specific needs of older adults. Some studies have found older adults have physically touched SARs even when they lack such capabilities [67], suggesting the potential in exploring physical touch beyond exercise environments [71] or pet therapy [78] to include tasks like ADL assistance. Older adults are more likely to use non-verbal utterances due to cognitive decline [41], presenting opportunities to improve accessibility by understanding the potential intent of these sounds in HRI.

An under explored interaction modality is the use of wearable sensors, namely for recognizing ADL related actions to monitor task progress. The following section of the literature review will explore wearable sensors used by SARs in greater depth to understand potential applications for ADL assistance, namely for assisting with the ADL of dressing.

How should SARs behave? SAR behavior that can adapt to user preferences and affective states has shown increased task performance [52] and compliance [69] among older adults. As developers seek to improve the social abilities of SARs, behaviors will need to focus on providing a personalized approach [110] while considering issues with respect to privacy, transparency, and user autonomy [37]. Further work is required to develop ethical frameworks specific to SARs with older adults to understand these concerns from a design perspective [111] similar to what has been done for telepresence robots [112].

An example requiring ethical consideration is the demand for a transparent approach to be taken to avoid developing deceptive or manipulative behaviors that will decrease long-term trust and efficacy of SARs, even if they gain short-term user compliance [113]. For emotional models, the relationship between cultural background and emotional expression [114] requires SAR behaviors to be sensitive to different cultural norms of intended users, and aim to create culture-neutral expressions [115] when possible. It is also critical to understand on an individual level how adults age, and what their needs and wants are with respect to SARs, as these can vary from one individual to another [116]. SAR behavioral models must account for such diversity in user abilities and aspirations [117].

Empathy is a underdeveloped promising strategy for SARs to use with older adults [110] which may be defined as "*The act of perceiving, understanding, experiencing, and responding to the emotional state and ideas of another person*" [118]. Empathy presents unique challenges as it requires integration of classification, adaptation, and emotional frameworks. For older adults, empathy has the potential to improve social stimulation and connection which is critical for applications that seek to decrease loneliness and depression [110]. Empathetic strategies specific to older adult mental health have already been an area of study in healthcare, and future SAR developments may use the outcomes from this research to design empathetic frameworks [119].

Classification, whether to enable empathy or other behavioral models, must consider the diverse number of different activities a user may complete. For ADL classification, observed activities may include both ADLs which the SAR has seen before during a training phase, as well as unknown ADLs. The last section of this literature review will discuss existing methods of ADL recognition, both broadly and more specifically for SARs, to understand key design features for improved performance in addition to limitations of current works.

2.2 Assistive Technologies for Dressing

This review was a contribution first published in [50]. Among ADLs, dressing is of particular importance as it enables self-expression, and the ability to dress oneself is associated with personal confidence [120]. Individuals with cognitive impairments face challenges in dressing associated with engagement in an activity and completion of the necessary activity steps, as they have reduced capacity to sustain attention on a sequential task and limited short term memory [121], [122]. To-date, systems used for providing dressing assistance include smart wardrobes [123]–[125], sensory devices for dressing state monitoring [126]–[128], and physically assistive technologies such as robot manipulators to dress a user [129]–[131].

2.2.1 Non-Contact Assistive Technologies for Dressing

Non-contact assistive technologies have been designed to: 1) provide clothing recommendations to a user by using smart wardrobes [123]–[125], or 2) track user dressing progress and provide feedback/corrective instructions via sensory devices to [126]–[128].

Smart Wardrobes

Smart wardrobe systems focus on the clothing selection task of getting dressed by identifying the dressing state through tracking when clothing is removed from its storage location [123] and in some cases also making clothing suggestions based on external information such as weather [124], [125]. For example, in [123] users with vision and/or hearing impairments were provided with customized devices which used a combination of visual (e.g., LEDs near clothing storage), vibration-based (e.g., wearables), and auditory (e.g., recorded voice over speakers) cues to help direct them to put on clothing items. Clothing items for a given outfit were tracked using signals from motion sensors on a customized wardrobe to identify when clothing items were selected by the user.

In [124], a smart wardrobe was developed for residents in assisted living centers. The wardrobe consisted of a wooden frame to hold the clothes, motion sensors used to detect clothing removal, LED strips aligned with clothing locations, and a Wi-Fi enabled tablet as a user interface displaying clothing suggestions. Rule based heuristics used indoor temperature from IoT sensors in the home, the weather forecast from an online database, and a user's event schedule using the Google Calendar API as input data to provide clothing recommendations. The tablet displayed clothing recommendations and the LEDS associated with the recommended clothes in the wardrobe were illuminated. Users could follow the suggestion on the tablet or browse for other options using its interface. User decisions were stored for future recommendations.

In [125], a smart wardrobe was developed to alleviate the burden of decision making required for clothing selection. RFID tags were attached to clothes in the wardrobe and were detected using an RFID reader placed above each clothing compartment. The user used a tablet to specify their mood and preferred colors. A matching algorithm used color matching and texture heuristics along with a random number generator to recommend clothing items. A limitation of all smart wardrobes is that they only assist with clothing selection rather than the entire dressing activity, due to their lack of both perceiving the user and interaction capabilities.

Sensory Devices

Devices used for tracking user dressing steps have focused on tracking clothing locations to provide feedback or corrective instructions using multiple sensors [126]–[128]. In [126], a customized dressing space was developed that included three RFID antennas positioned in front of a mirror, on a clothing hanger, and on a wardrobe to detect embedded RFID tags on clothing. Dressing errors (e.g., forgetting to put on a clothing item, putting on clothing items in an incorrect order) were identified using a Layered Hidden Markov Model based on the time series location data.

In [127], a camera was positioned facing a user to detect dressing errors classified as temporal (wrong order), relational (wrong orientation), or spatial (partially worn). Support vector machines (SVMs) were developed and trained on color histograms and texture information using a gray-level co-occurrence matrix and local binary patterns to identify the clothing item on the user.

In [128], a dressing state evaluation system was developed using: 1) a tablet that visually displayed the dressing steps to be completed in list format, 2) motion sensors adjacent to clothing item storage locations for monitoring when a user selects the item, 3) skin conductance wearable sensors for monitoring stress, 4) RFID readers and tags for clothing localization, and 5) a RGB camera and custom fiducial markers for determining clothing orientation. A series of rules were used for each dressing step using the location of all fiducial markers to determine the specific dressing classes such as "*both arms of shirt worn*" and "*partial dressing (incomplete)*". The motion sensors were used to detect if clothing was removed before it was time to put it on. Furthermore, the system could determine when a user was “stuck” by combining recent task progress with skin conductance-based measurements to trigger audio appropriate instructions.

The sensory devices that have been developed for dressing assistance have shown success in tracking dressing steps based on clothing item location using cameras and RFID tags. However, these solutions are limited in their ability to identify important dressing actions such as buttoning/zipping up as they classify dressing steps using only clothing locations. The wearable sensory system directly tracks user motion to enable classification of all dressing steps including those that do not require a change in clothing location/position. This thesis builds on existing work in smart clothing for motion tracking as outlined below.

2.2.2 Assistive Technologies for Dressing using Robotics

Robotic technologies used for the dressing ADL have either consisted of robot manipulator arms to physically help with dressing [129]–[131] or socially assistive robots providing prompts and feedback to guide users through the dressing steps [132]–[134].

The Baxter robot has been used for physical dressing assistance using user joint positions obtained using 1) RGB-D cameras with skeleton tracking [129], 2) RGB-D cameras with occluded joint position estimation based on positions of visible joints [130], and 3) built-in torque sensors with probabilistic human joint estimation [131]. In [129], user position tracked using an RGB-D camera was used to generator manipulator velocities based on the shortest path or repositioning requests when the user was out of reach. This approach was limited by camera view occlusion from the clothing and Baxter’s manipulators. In [130], a recurrent neural network (RNN) was used to estimate elbow joint locations when the elbows were occluded by upper body clothing during dressing. A regression trees approach determined which visible joint locations, such as shoulders and hips, to use as features for input to the RNN. In [131], Baxter’s built-in force-torque sensors with a hierarchical multitask controller using a probabilistic model were used to determine manipulator trajectories. The controller was trained to minimize the force between the user and the robot, eliminating the requirement for a clear view of the user.

In [132], a sensor array was integrated into a smart collared shirt for monitoring dressing states in order for the Pepper robot to use to provide social dressing assistance. The sensor array included: 1) a front mounted IR LED for detecting orientation between front/back of the shirt, 2) contact switches on the buttons for determining fastening, and 3) arm sleeve and back mounted capacitive switches for detecting contact with human skin for determining partially worn states. User dressing states identified using the combination of these sensors included correctly worn, partially worn, backwards, or inverted.

In [133], [134], the socially assistive robot Leia was introduced to provide both verbal and visual clothing recommendations using a developed app displayed on a tablet adjacent to the robot. Users were guided through a clothing selection process based on weather, user preferences and dress code, and activity plan information including if they will be active or outdoors. The robot then used a multinomial logistic regression (MLR) method for providing an ordered list of suggestions. while learning recommendations through user interactions. Depending on user clothing item selection, user preferences for specific items were updated using stochastic gradient descent.

Robot manipulators have focused on physically dressing people, rather than encouraging them to self-dress, which can result in individuals losing their own ability to dress themselves [10]. In general, social dressing assistance has been found to improve the independence and functional performance of cognitively impaired people by empowering them to complete activities using their own abilities [135]. Socially assistive robots can provide prompts and progress feedback to users, in order to motivate them to dress themselves. However, to-date, socially assistive robots have only been used for the clothing selection task and not the dressing task itself.

2.3 ADL Recognition for SARs

This review was first published in [51]. Existing work in human activity recognition (HAR) utilizes a variety of methods to improve accuracy, enable unsupervised learning, and adapt to specific scenarios and users. This section provides a detailed discussion on: 1) multimodal DL networks for HAR, 2) embedding of feature spaces for HARs, and 3) HAR for social robots.

2.3.1 Multimodal DL Networks for HAR

Recent HAR research has focused on using data specific operations such as graphical convolution networks (GCNs) [136], [137] and learned fusion techniques [138] to combine multimodal inputs in order to extract complimentary features. These methods have combined a variety of inputs including human skeleton pose information [136], [137], RGB video [136]–[138], and motion information between frames [138].

In [136], both RGB video and 3D poses were used to extract visual features for spatial embedding and pose driven attention using a Video-Pose Network consisting of a combination of GCNs and spatio-temporal convolution networks in order to classify indoor activities, such as putting on headphones and clapping for monitoring of human behavior. End-to-end training with 3D ConvNet using a regularized loss term combining cross-entropy, embedding loss, and an attention regularizer resulted in significant improvements in classification accuracy for subtle actions such as reading compared to single mode networks using only RGB video streams. In [137], the architecture proposed in [136] was further extended through the incorporation of two separate distillation training sessions. Distillation transferred knowledge of pose to the feature extraction layers for improved model speed using only the RGB video input.

In [138], an RGB video stream and a motion information stream obtained from persistence of appearance (PA) were both used by a spatio-temporal convolutional neural network (CNN) with modality specific attention and late fusion for ADL classification. Training was accomplished using classification loss to learn consensus attention between the two modalities. Testing on segmented/unsegmented RGB video data of users performing ADLs, such as eating with a fork, showed improved accuracy over using a single modality.

2.3.2 Embeddings of Feature Spaces for HAR

Embeddings of feature spaces are used to learn low-dimensional vector representations of ADLs to reduce the dimensionality of categorical information. They are used for data visualization [139], and classification [136], [140]–[142].

In [139], accelerometer and gyroscope sensory data from users completing ADLs was reduced to an embedding vector of activity features. An Autoencoder based on a Long-Short Term Memory Recursive Neural Network architecture was trained to reduce and reconstruct the sensory data for training the embedding. Temporal features were embedded using a sequence of recursive convolutions for activities of variable lengths. 2D Visualization of the embeddings based on stochastic neighbor embedding (t-SNE) [143] showed that the embedded features had improved inter-class separation compared to handcrafted features for the same data.

In [136], feature embeddings were used for multimodal fusion to improve the classification accuracy of the Video-Pose Network. An intermediate spatial embedding space was developed by combining RGB video visual features and pose spatial features. Embedding loss was added to the training loss function which improved inter-class separation.

In [140], data from inertial measurement units (IMUs) was mapped to a feature embedding vector to enable classification of sparsely labeled data. Feature embeddings were derived from the temporal input using CNNs and contrastive learning. Using the feature embeddings on partially labeled data for movement activities showed classification accuracy improvements over existing conventional autoencoders. In [141], a self-attention based approach was developed using a hierarchical window encoder (HWE) trained on temporal activity data using reconstruction loss to create feature embeddings. These embeddings were used for both classification training with unlabeled data and open-set recognition to identify unknown activities. A dense neural network with non-linear activations encoded and decoded the embedding features using an autoencoder architecture with an activity embedding vector. Training results showed improvements in closed-set classification over conventional autoencoders. Testing with unknown activities confirmed the ability of the network to identify such activities.

Robotic object manipulation has also used embedding vectors for classification. For example, in [142], point clouds of objects, natural language instructions, and robot manipulation trajectories were embedded in a common embedding space using linear deep neural layers with non-linear activations. The feature embeddings were used to select a new manipulation trajectory based on the embedding of an object-instruction pairing. Results showed improved accuracy and speed compared to embedding models using the same approach with larger and more complex embedding spaces

2.3.3 HAR for Social Robots

HAR has been used by social robots in human-robot interactions for numerous applications ranging from playing games to companionship [144]. A handful of SARs have been used to classify and track users performing ADLs using unimodal RGB video [52], [58], unimodal pose data from a depth sensor [39], [145], and multimodal data from RGB video and object-based sensors [52]. These activity tracking systems mainly use visual and depth data [39], [52], [58], [145] or natural language [39] classifiers, and heuristic rules [52], [58] to monitor and provide feedback to users via SARs.

In [52], the human-like Brian robot was used to facilitate meal eating of older adults. A sensor suite was used consisting of Wii motes for tracking custom IR utensils, a Kinect sensor for detecting user engagement from pose, and a meal tray with embedded load cells. The user and activity state, determined using the sensory information and Haar feature-based cascade classifiers and decision rules, and the robot state, based on task progression history, were used by a finite state machine (FSM) to determine the robot's assistive behavior. In [58], the Bandit robot was used to engage older adults in workout, imitation, and memory games. User hand and elbow joint positions were classified using an image segmentation algorithm and heuristic exercise rules. They were then used by an FSM to provide verbal praise by the robot for successful actions and corrections for unsuccessful actions.

In [145], the Leia robot was used to guide users in upper body exercises. An RGB-D sensor extracted user poses and a K-nearest neighbors classifier was used to classify these poses to determine exercise completion. An FSM was used to determine robot behaviors based on the exercise goal and user state. In [39], the human-like Casper robot learned to assist users in the ADL of making a cup of tea. The robot used a combination of Learning from Demonstration and reinforcement learning to determine its task-related assistive behaviors based on user cognitive functioning and activity states. Learning of task-related behaviors was based on demonstrators' speech using the onboard microphone and IBM Watson Speech-to-Text API [83] as well as gestures obtained by a depth camera and tracked using OpenNI [146].

2.4 Chapter Summary

This chapter began with a review of existing social robots for assisting older adults. Design features of appearance, interaction modes, and behaviors were explored in terms of their use in existing robots and their effect on users as determined from a variety of studies. Additionally, SAR overall autonomy was found to be a barrier to long-term deployment as existing robots are designed for one or a few tasks in a structured environment. Identifying dressing as a potential ADL for a novel wearable interaction mode, existing technology for dressing and wearables for SARs were then discussed. Lastly, motivated by the limitations of existing SAR behaviors (only designed for a single ADL) and SAR autonomy presenting a barrier to long-term use, existing research on activity recognition was reviewed. Findings from each section of this review motivate the development of new technologies that increase SAR ability and intelligence for providing ADL assistance.

CHAPTER 3

3. A SOCIAL ROBOT AND WEARABLE SYSTEM FOR DRESSING ASSISTANCE

The novel Socially Assistive Robot-Wearable system architecture is presented in Figure 1. The *Strain Sensor Smart Clothing* is used to obtain resistance signals for tracking user actions during the dressing activity. These signals are used by the *Joint Angle Mapping Model* to estimate user joint angles. Joint angles are then provided to the *Dressing Step Classifier* for DL classification of dressing actions. Dressing actions are used by the *Robot Adaptive Behavior Deliberation* module to determine appropriate assistive behaviors for the socially assistive robot Leia to display using a combination of verbal and non-verbal communication modes via its low-level controllers in the *Actuation* module. This work was first published in [50] and herein this thesis focuses on the developments as they relate to the socially assistive robot and system integration. Namely, the *Robot Adaptive Behavior Deliberation* module and resulting robot behaviors are discussed.

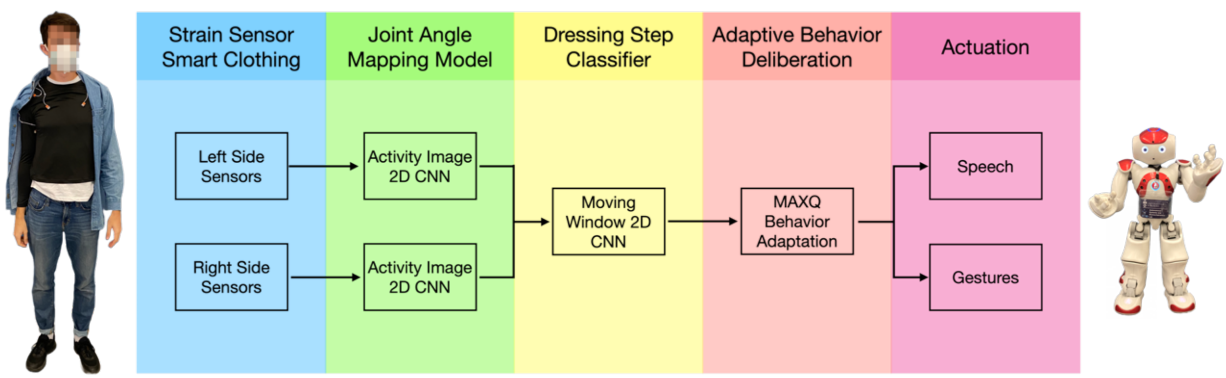


Figure 1. Socially assistive robot-wearable system for dressing assistance proposed architecture [50].

3.1 Robot Adaptive Behavior Deliberation

The robot adaptive behavior deliberation module uses a MAXQ reinforcement learning hierarchical method [147] to determine Leia's assistive behavior based on the user state and past user actions. MAXQ is used as it encompasses: 1) temporal abstraction to allow for variable dressing step completion times, 2) state abstraction for reducing the user dressing state space at each level to only relevant variables, and 3) subtask abstraction for grouping similar actions (e.g., *Put Through* groups "left arm through", "right arm through", and "head through" steps) and further reduces the state space. The MAXQ hierarchical task graph showing the Markov decision process (MDP) decomposition for the *Root Task* of dressing is shown in Figure 2.

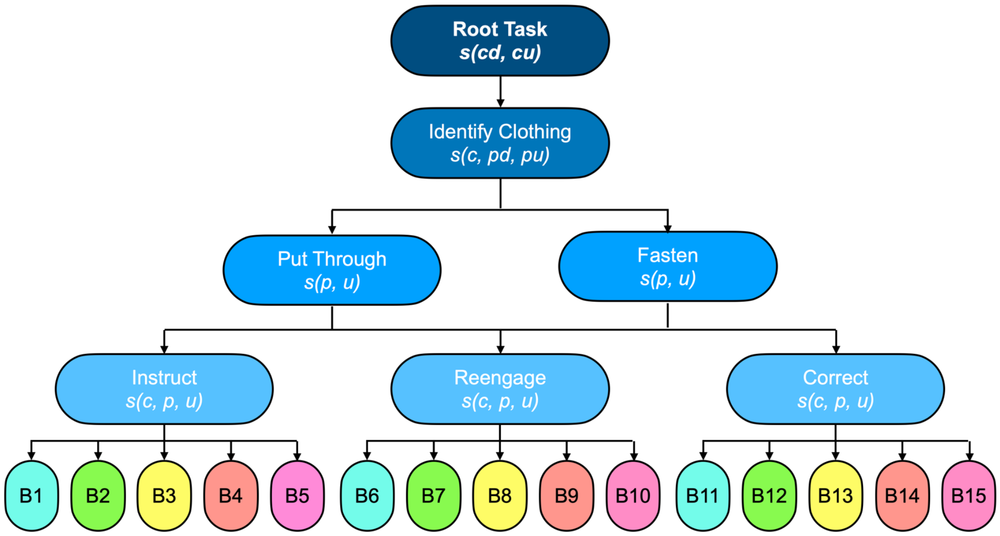


Figure 2. MAXQ socially assistive robot-wearable system dressing task graph [50].

For each level of the MAXQ graph, the associated subtasks have a defined set of relevant state variables *s* and a terminal condition *T,* where all variables are expressed using one-hot encoding. State variables and their descriptions are presented in Table 1. For each level of the hierarchy, *s* and *T* are defined as: 1) *Root Task, s(cd, cu)*, *T: cu = cd*; 2) *Identify Clothing, s(c, pd, pu)*, *T: pu = pd*, 3) *Put Through/Fasten, s(p, u)*, *T: u = p*, and 4) *Instruct/Reengage/Correct, s(c, p, u)*, which terminates on completion of a primitive action (*B1-B15*).

Table 2. Behavior utterance and gesture examples [50].

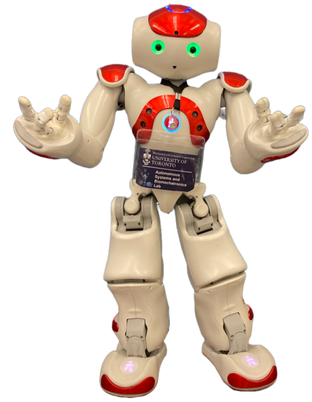
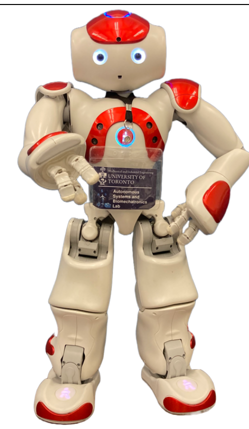
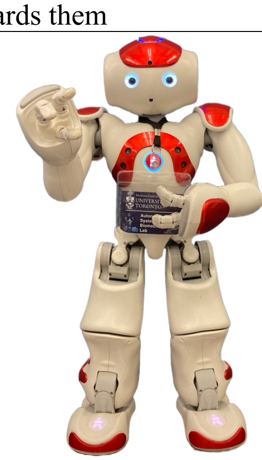
State Variable	Description	Example
cd	the set of clothing items in an outfit	{"t-shirt"; "button up shirt"; "jacket"}
cu	the set of clothing items on the user	{"t-shirt"}
c	the current clothing item for assistance	"button up shirt"
pd	the set of dressing steps for *c*	{"left arm through"; "right arm through"; "button up"}
pu	the set of dressing steps completed by the user	{"left arm through"; "right arm through"}
p	the current dressing step for assistance	"button up"
u	the user state expressed as the most recently completed dressing step	"right arm through"

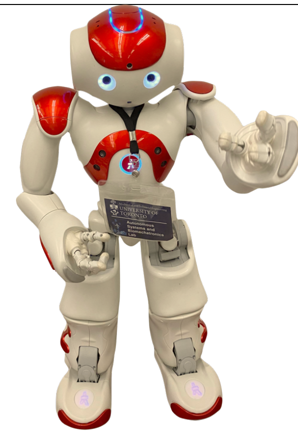
Primitive actions are completed by Leia using a specific behavior expressed using both verbal utterances and non-verbal gestures in the form of body language to emphasize spoken ideas [148]. Primitive actions *Instruct (B1-B5)*, *Reengage (B6-B10)*, and *Correct (B11-B15)* are selected based on the previous user step u and current dressing step p. When $u = p$, the user completed the desired dressing step and Leia *instructs* the next step. When $u = ia$ (representing inaction), Leia *reengages* the user to complete the current step p. When $u \neq p$ *or* ia, the user has completed an incorrect dressing step and Leia helps correct this error by asking the user do undo the previous step.

Robot behaviors for primitive actions (B1-B15) were selected based on different compliance gaining behavior (CGB) strategies [149], [150] including logic, emotion, direct request, cooperative, and motivation. Namely, the 5 potential robot behaviors for Instruct (B1-B5), Reengage (B6-B10), and Correct (B11-B15) include one of these CGB strategies. Logic and emotion strategies are based on HRI research that has shown these two CGBs are most effective for social robots in persuading users [150]. The remaining three strategies are based on the clinical experience of caregivers assisting older adults with the dressing task [135] and guidelines provided by the Alzheimer's Society for effective dressing assistance which focus on clear communication, creating a sense of teamwork, and providing consistent positive verbal encouragement [151]. The combination of HRI research and clinical experience as selection criteria for the robot behaviors was used to include behaviors that consider the robot form and function. Since dressing task assistance is a new domain for SARs, behaviors are not ranked or given any initial preference during behavior deliberation. Examples of each behavior strategy for the Instruct behavior (B1-B5), the Reengage behavior (B6-B10), and the Correct behavior (B11-B15) are presented in Table 2. For a specific behavior strategy type, the robot's speech differs, however the gesture is the same.

Table 3. Behavior utterance and gesture examples [50].

Behavior Strategy	Utterances	Gestures
Logic	For B1 (instruct): "My sensors tell me it is time to button up your dress shirt." For B6 (reengage): "My sensors tell me you are disengaged with the task. Please finish buttoning up your dress shirt." For B11 (correct): "My sensors tell me you made a mistake while	

Behavior Strategy	Utterances	Gestures
	doing up your buttons. Please undo the buttons on your dress shirt.”	References self
Emotion (Happy)	For B2 (instruct): “It would make me happy if you buttoned up your dress shirt.” For B7 (reengage): “It would make me happy if you were reengaged in the task and buttoned up your dress shirt.” For B12 (correct): “There was a mistake while doing up your buttons. It would make me happy if you undid the buttons on your dress shirt.”	 Expansive and open gestures and green eye color for positive emotion
Direct Request	For B3 (instruct): “Please button up your dress shirt.” For B8 (reengage): “Please reengage in the task and button up your dress shirt.” For B13 (correct): “There was a mistake doing up your buttons. Please undo the buttons on your dress shirt.”	 References user by pointing towards them
Cooperate	For B4 (instruct): “Let’s work together to button up your dress shirt.” For B9 (reengage): “Let’s reengage in the task and button up your dress shirt.” For B14 (correct): “There was a mistake while doing up your buttons. Let’s undo the buttons on your dress shirt.”	 References user and self

Behavior Strategy	Utterances	Gestures
Motivate	For B5 (instruct): "Please button up your dress shirt. You can do it!" For B10 (reengage): "Reengage in the task and button up your dress shirt. I know you can do it." For B15 (correct): "You made a mistake while doing up your buttons, but that's okay. To fix it, please undo the buttons on your dress shirt."	 References user while performing head nodding action to show support

3.2 Performance Testing

Performance testing was done on the different modules of the architecture and the overall robot-wearable system to evaluate reliability for completing designed functionality, namely behavior adaptation, dressing step classification, and upper body dressing assistance.

3.2.1 Robot Adaptive Behavior Deliberation Cumulative Reward and Convergence

The performance of the *Behavior Adaptation Model* was evaluated based on: 1) convergence of the MAXQ learning model in learning an optimal policy for user dressing, and 2) online task performance measured by cumulative reward per dressing iteration. A dressing iteration is defined as the entire action set to put on all clothing items in an upper were modeled with varying actions and preferences. For each user, their compliance, engagement, and frequency of mistakes were defined using probability rates dependent on their preferred robot behavior. Rewards at each level of the task hierarchy were given as follows: *Root Task*: ±5, *Identify Clothing*: ±3, *Put Through/Fasten*: ±1, *Instruct/Reengage/Correct (Success)*: 0, *Instruct/Reengage/Correct (Failure)*: -1. Offline training was performed with *U1* to evaluate convergence of MAXQ model training, while *U2-U5* were used for online testing.

The values of the MAXQ learning parameters were set to: 1) learning rate α = 0.01; 2) exploration/exploitation trade-off $\varepsilon = 1 - \log(1 + \lambda i)$ for $\varepsilon > \varepsilon_{min}$, where $\varepsilon_{min} = 0.05$ to maintain the ability to adapt to changes in user preferences over time; and 3) rate of logarithmic change $\lambda = 0.1$.

The MAXQ model was trained for 10,000 iterations on *U1*. It converged to the optimal policy within 10 iterations. Testing with U2-U5 to determine the model's online adaptation capabilities to new users and preferences is shown in Figure 3. A higher cumulative reward indicates fewer number of steps taken by the user to dress. A hierarchical system was also designed for comparison purposes, which uses the task structure and state transitions of the MAXQ, however, selects a random behavior type instead of learning behavior preferences.

As can be seen in Figure 3, the overall cumulative reward for MAXQ increases in all cases compared to the initial dressing task iteration as Leia learns behavior preferences. Acceptable performance is achieved for the MAXQ method as behavior selection converges in terms of cumulative reward and is optimized for each user. Furthermore, the cumulative reward for MAXQ compared to the hierarchical system without behavior preference learning shows improved task performance when behavior learning is used for users with weak preferences (e.g., U2 and U3) and strong preferences (e.g., U4 and U5).

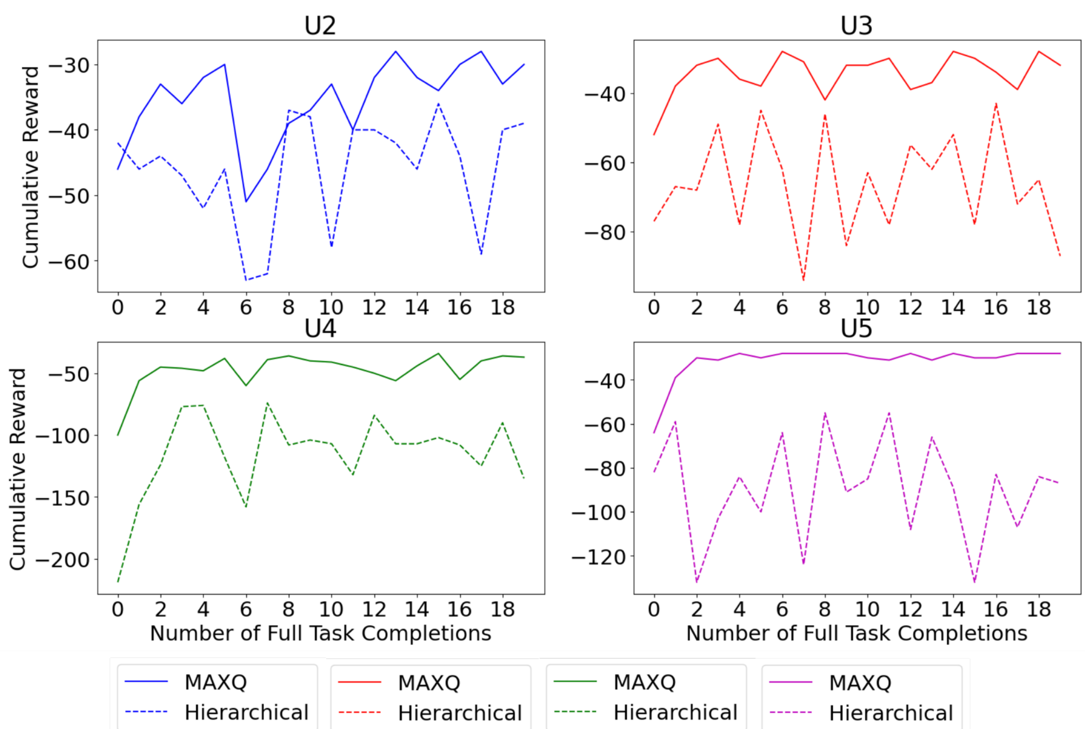


Figure 3. Cumulative reward of users for measuring task performance given by MAXQ and a hierarchical system without behavior preference learning [50].

3.2.2 Dressing Step Classifier Accuracy

The classification accuracy of the *Dressing Step Classifier* was obtained using the test dataset collected from a cognitively healthy user performing dressing in upper body outfits that used all dressing actions (e.g., t-shirt, button up shirt, and jacket). The labelled dressing step test set contained 122 evenly distributed samples of 25 joint angle arrays. Classification accuracy is shown as a confusion matrix in Figure 4 with an overall classification accuracy of 95.1%. The "head through" class had a 96% accuracy and the lowest classification accuracy for all dressing steps was 91% for "left arm through". The success rate of 95.1% achieved by the smart clothing for dressing step classification is higher than those achieved by other smart clothing PZT sensors. For example, in [152], four PZT sensors integrated in a loose-fitting jacket had an accuracy of 90.9% when classifying 5 postures and activities from sensor signals, including standing, running, sitting, lying, and walking.

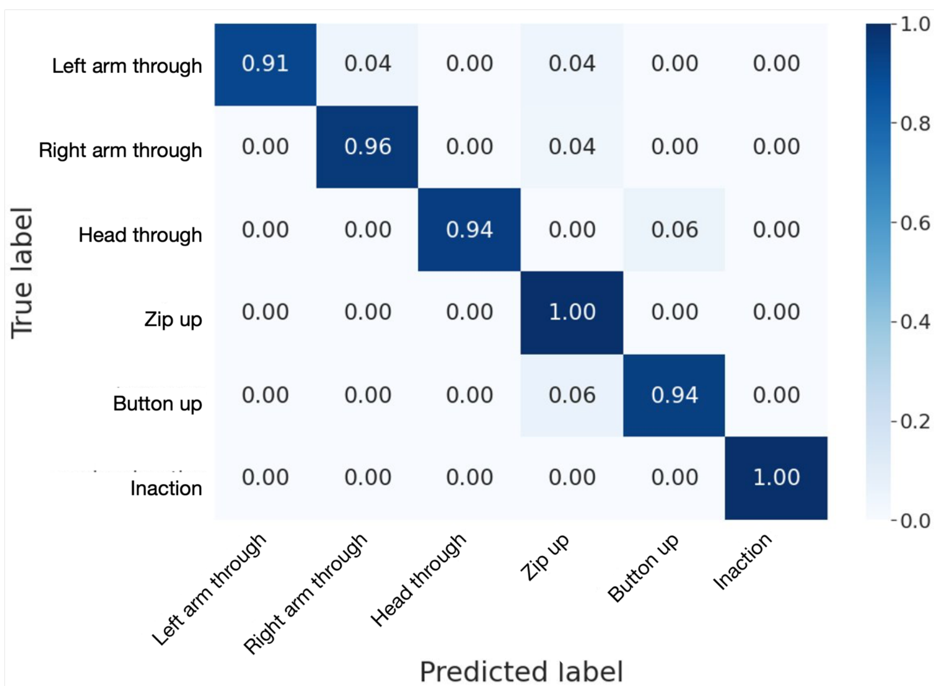


Figure 4. Dressing step classifier test accuracy in confusion matrix [50].

3.3 Demonstration Study

A robot demonstration study, initially published in [50], was conducted during technology conferences, where the capabilities of the socially assistive robot-wearable system were presented to assistive technology, robotics, medical, and healthcare researchers and entrepreneurs to gain their insights on the overall system design and functionality. The robot demonstration study consisted of one researcher demonstrating putting on and buttoning a dress shirt with social assistance from Leia, while wearing the smart sensors, Figure 5. The study was approved by the University of Toronto's Health Sciences Research Ethics Board (REB), protocol #43592.

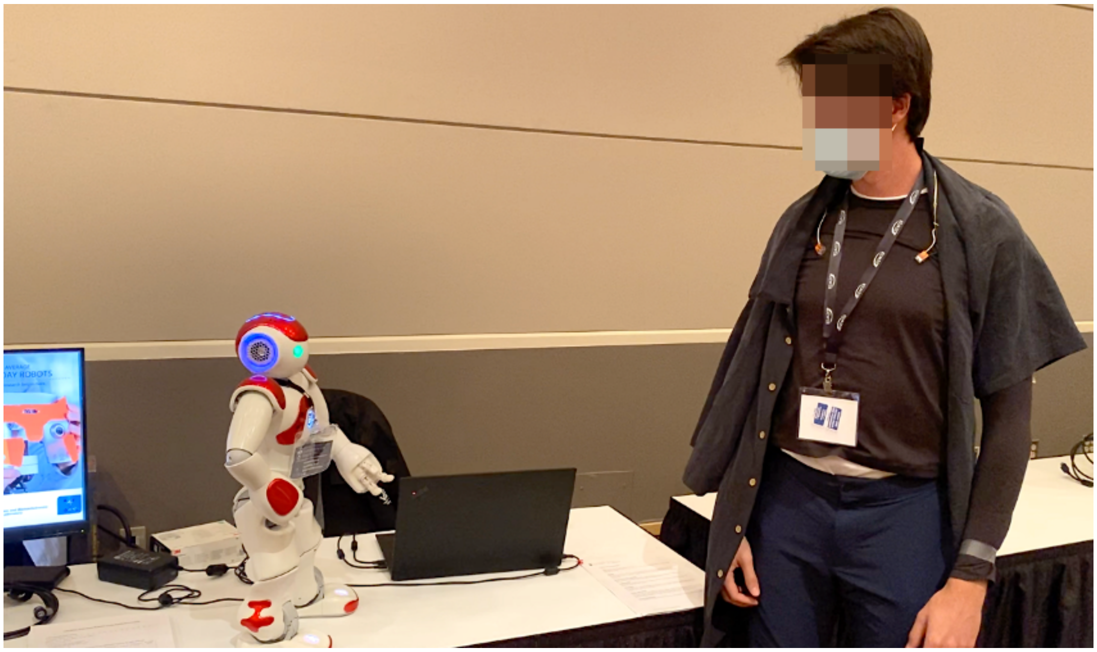


Figure 5. Demonstration setup at technology conference [50].

3.3.1 Participants

Participants were recruited during the exhibitions for two technology conferences. The first conference was on assistive technology for older adults and consisted of experts in assistive technology and device development, healthcare and medical professionals and researchers. The second conference was on AI technology and consisted of researchers and engineers working at tech companies. Approximately 100 participants observed the robot demo, and thirty-two participants filled in the questionnaire. The participants ranged in age from 20-60 years old, with the majority being in the 20-29 age group.

3.3.2 Procedure

The demonstration consisted of Leia introducing itself and assisting the demonstrator in putting on a dress shirt. The robot randomly used two of its five behavior strategies (outlined in Table 1 of Section 3.4) to guide the demonstrator in the dressing task. A video of the demonstration is provided here: https://youtu.be/Klf0O-OSEik

3.3.3 Measures

A 5-point Likert Questionnaire (1-Strongly Disagree 3-Neutral, 5-Strongly Agree) was administered to participants after the demonstration was completed, Table 3. The questions were adapted from the Almere model [153]. They included statements on the following attributes: A1 – *perceived usefulness* (Q1-3), A2 – *perceived ease of use* (Q4-6), A3 – *attitude towards* (Q7-10), and A4 – *satisfaction with socially assistive robot-wearable system* (Q11-12). It was chosen to include a *satisfaction* attribute to measure stakeholder satisfaction of the system to assist the target population with dressing, namely individuals with cognitive impairments.

Table 4. Study questionnaire organized by attribute with descriptive statistics [50].

Construct	Question	Median ($\tilde{x}$)	IQR	Min	Max
Perceived Usefulness	**Q1.** Using the social robot-wearable system would make it easier to dress	4	1	1	5
	Q2. I would find the social robot-wearable system useful for dressing	4	2	1	5
	Q3. Using the robot system would increase dressing performance	4	1	1	5
Perceived Ease of Use	**Q4.** I find the social robot-wearable system easy to use	4	2	1	5
	Q5. The interaction with the social robot-wearable system is clear and understandable	5	1	2	5
	Q6. The social robot-wearable system is easier to use than other dressing aids	3	1	1	5
Attitude Towards	**Q7.** I think the social robot-wearable system can adapt to what is needed for the dressing task	4	1	1	5
	Q8*. If I (or someone I know) used the social robot-wearable system, I (they) would be afraid to make mistakes	2	1	1	4
	Q9. The social robot-wearable system looks fun to use	5	1	2	5
	Q10. The wearable sensors appear to be comfortable to wear	3	2	1	5
Satisfaction	**Q11.** I would recommend the social robot-wearable system to a friend	3	1.25	2	5
	Q12. I think it is a good idea to have the social robot-wearable system	4	1	2	5

* Negatively worded question

3.3.4 Results

A Shapiro-Wilk test confirmed the data was non-normal (p<0.05). Statistical analysis was conducted using the non-parametric Mann-Whitney U tests (MWU). Overall study results showed positive ratings for the majority of the questions on perceptions of the socially assistive robot-wearable system, Table 3 and Figure 6. A1 had consistent positive ratings and internal consistency between question ratings ($\tilde{x}$=4 for Q1- 3). Attributes A2, A3, and A4 showed greater intra-attribute question rating variation with some strongly-agree ratings ($\tilde{x}$=5 for Q5 and Q9), some agree ratings ($\tilde{x}$=4 for Q7 and Q12), and neutral ratings ($\tilde{x}$=3 for Q6, Q10, and Q11). These ratings showed that participants noted the system was easy to understand and use, as well as being fun and a good idea. Neutral ratings were given for questions on comparisons with other tech, comfort, and recommendation to a friend. It is important to note that the negatively worded Q8 ($\tilde{x}$=2) showed that stakeholders believed that users would be able to use the robot-wearable system and would not be afraid in using it for the dressing task.

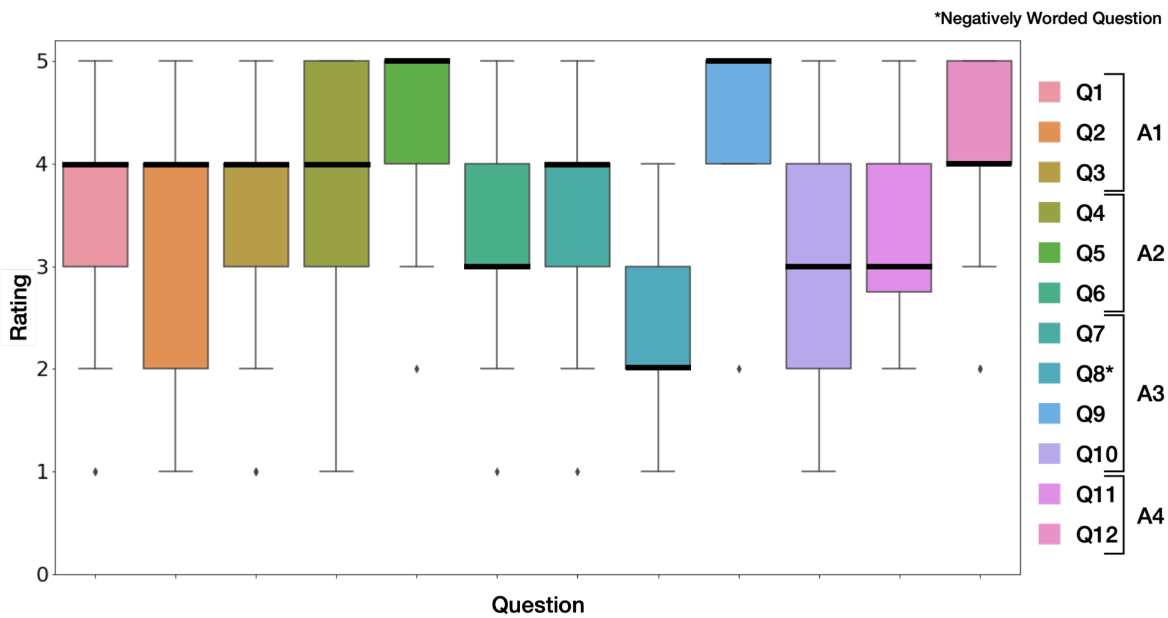


Figure 6. Box and whisker plot of questionnaire ratings for each question for all participants. Median is represented as bold lines IQR is represented by the boxes [50].

Gender

Twenty-seven of the participants provided their gender as either Female (*n*=16); Male (*n*=10); or Other (*n*=1). The descriptive statistics are shown in Figure 7. MWU tests were conducted to determine if there were any statistically significant differences between the ratings from male and female participants for all questions. A statistically significant difference was found for Q7- "*I think the social robot-wearable system can adapt to what is needed for the dressing task*", between males ($\tilde{x}$=4, *IQR*=0) and females ($\tilde{x}$=4, *IQR*=1), MWU test: *Z*=2.319, *p*=0.020. There was also a statistically significant difference for Q10- "*The wearable sensors appear to be comfortable to wear*", between males ($\tilde{x}$=4, *IQR*=1.25) and females ($\tilde{x}$=3, *IQR*=1.75), MWU test: *Z*=2.013, *p*=0.044. Namely, males provided more positive ratings than females for these questions.

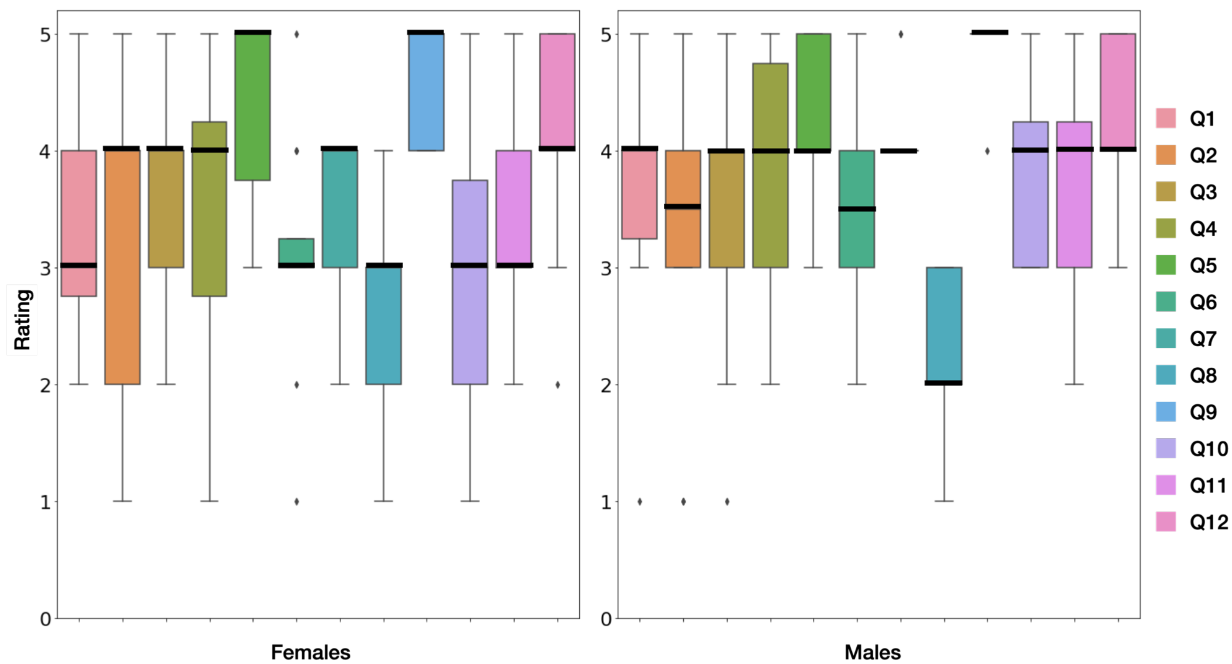


Figure 7. Box and whisker plot of questionnaire ratings between females and males. Median is represented as bold lines IQR is represented by the boxes [50].

Participant Feedback

Participants also provided comments with respect to their overall experience. Similar comments were grouped together into the following categories: 1) system appearance such as "*I like the make/external body of the robot, makes it very approachable.*"; 2) recommendations for future development including "*Wonderful. Look into patentable systems and commercialization - cost to make the robot.*"; and 3) appraisal for potential users such as "*The robot is super cute and I see how this may be helpful for those who may be cognitively vulnerable and may not know how to do simple daily tasks.*"

3.3.5 Discussions

In general, stakeholders noted the usefulness of the prototype and confirmed its use for the intended population with overall positive ratings in the questionnaire. This was supported by comments such as "*I would recommend it to a friend with mobility or cognitive difficulties.*" Statistically significant differences between males and females were found for two (Q7 and Q10) of the four questions in A3 with males having higher ratings. The tendency for males to have a more positive attitude towards socially assistive robots has also been found in other care application studies [154], [155]. Gender differences in attitudes towards wearable technology for health have been application dependent. For example, males have shown higher intent to use smartwatches while having lower intent to use skin protection wearables than females [156], [157]. This application-based variation makes the smart clothing for dressing a unique contribution in this field.

3.3.6 Considerations and Limitations

The demonstration study took place in exhibition areas with crowds of people walking by and stopping to observe. Due to this scenario setup, it was not feasible to have everyone fill in the questionnaires, however, many verbal positive responses were received related to system usefulness and ease of use. In general, people were engaged in the demonstration. They reacted to the robot's behaviors by waving at Leia, speaking to the robot (e.g., "*hello*", "*what's your name*?", "*hi Leia*"), laughing, nodding, and smiling indicating a positive experience and sustained attention. While people did not directly interact with the developed system, the robot's verbal and nonverbal communication modes made the interaction easy to understand and the robot's intent was clear as mentioned by many who viewed the demonstration. Social robot demonstrations have been used to display robot capabilities in the early stages of development for acquiring feedback from potential users [158] and stakeholders such as caregivers [159] and therapists [160]. The robot demonstration study took place at conferences which provided access to a diverse group of stakeholders in a single location.

Participant pool bias may be present in the feedback received from relevant stakeholders given the conferences were focused on various types of technology. However, stakeholder familiarity with different types of technology reduces the risk of new information bias, as it has been shown that respondents unfamiliar with new technology can actually rate novel systems more positively [161]. Completion of the questionnaire required individuals to approach the demonstration. This may result in response bias where more extreme responses, both positive and negative, are received from respondents who actively seek to provide feedback [162]. However, for the questionnaire results, even though some extremely positive and negative responses were observed, the overall median ratings showed that the majority of responses were moderate, suggesting response bias did not have a significant effect on results.

3.4 Chapter Summary

This chapter presented the first socially assistive robot-wearable sensors system to provide dressing assistance through social HRI. A novel robot-wearable architecture has been developed to recognize and classify user dressing actions and provide personalized prompts and feedback. The assistive robot uses a MAXQ hierarchical learning method to learn appropriate assistive behaviors to guide the user through a sequence of upper body dressing steps. Experiments were conducted that validate the performance of the robot-wearable system to identify and effectively respond to a variety of user states and dressing step actions. Furthermore, a robot demonstration study with stakeholders found that overall, they had positive perceptions and attitudes towards the socially assistive robot-wearable system, in particular with respect to its usefulness with the intended user population.

CHAPTER 4

4. ADL RECOGNITION FOR SARs

This work was initially published in [51]. The objective of the deep learning human activity recognition and classification architecture is to identify ADL classes for SARs to assist with and monitor performance during activity completion.

4.1 Network Architecture

The overall proposed architecture is presented in Figure 8. Environment, action, and object information is obtained from RGB-D videos. These videos are separated into multiple inputs to extract pertinent features. Namely, a downsized RGB video is obtained from the RGB channels and used by the *Video Backbone Network* to obtain a combination of scene and motion features. The 3D pose of the user is simultaneously obtained from the RGB video and depth streams and used by the *Pose Backbone Network* to obtain 3D user motion features independent of the scene context. Single RGB images are also extracted and used by the *Object Detection Network* to obtain semantic features for objects used in performing the ADLs.

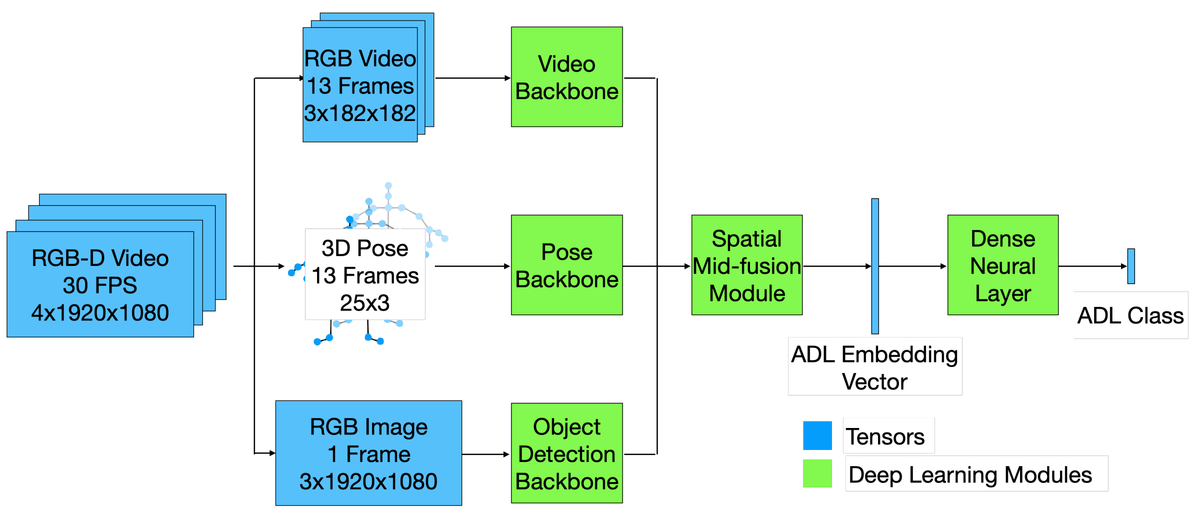


Figure 8 - Proposed DL ADL recognition and classification architecture [51].

The extracted feature set containing scene, motion, and semantic features from these backbone networks is then utilized by the *Spatial Mid-Fusion Module* to reshape and spatially scale the features for alignment before concatenation. This module condenses the features to a one-dimensional ADL embedding vector. The embedding vector is used by the *Dense Neural Layer* to determine the appropriate ADL class. The following subsections discuss these modules in more details.

4.1.1 Video Backbone Network

The objective of the *Video Backbone Network* is to extract scene and motion features, Figure 9. The network takes as input a sequence of 13 video frames of size *182 x 182* pixels by down sampling and cropping from the 30 fps RGB video stream. The X3D small network [163] is adapted herein as the feature selection method as it is a deep network designed for optimized video feature extraction. The X3D small network progressively expands spatial and temporal convolutional layers based on the ResNet architecture [164] in dimensions of temporal duration, frame rate, spatial resolution, width, bottleneck width, and depth in order to iteratively add model depth to achieve accuracy while decreasing complexity [163].

The layers of the X3D small model used herein are ResNet Stem which consists of a 2D spatial convolution for spatial feature extraction, a 1D temporal convolution for temporal feature extraction, batch normalization [165] to increase training speed and model generalizability, and rectified linear units (ReLU) activation [166] to introduce non-linearities while avoiding vanishing or exploding gradients. Four successive ResNet Stages, each with varying branch quantities and compositions follow as demonstrated in Figure 9. The output of the last ResNet Stage of X3D small is *13 x 192 x 6 x 6* (time, channels, feature grids) where each *6 x 6* video feature grid has inherent spatial understanding relative to the initial video.

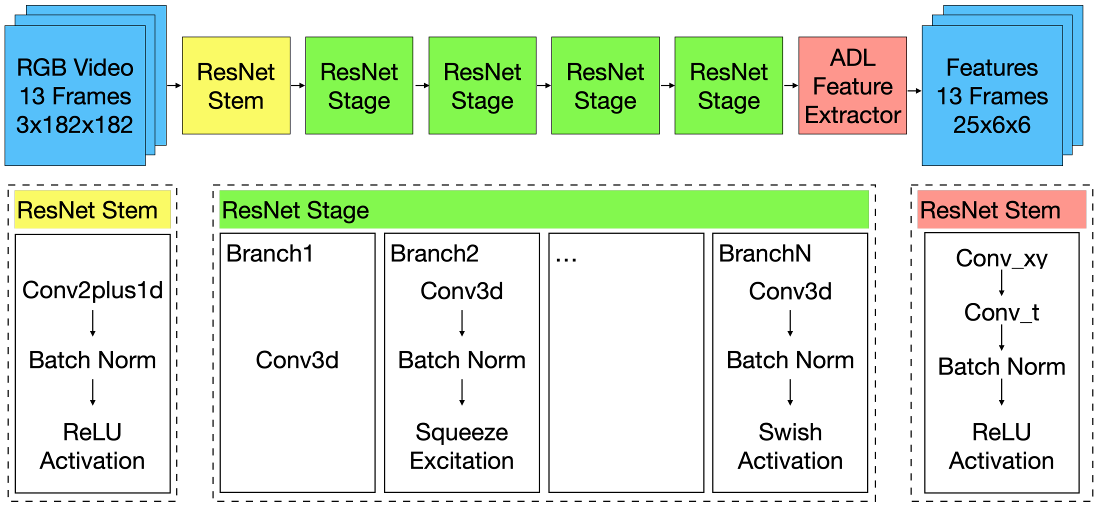


Figure 9 - Video backbone network architecture [51].

An ADL Feature Extractor was designed to select the most significant video features for classifying ADLs using the general extracted features from the X3D layers as input. The ADL Feature Extractor uses a spatial convolution for geometric features, a temporal convolution for motion features, batch normalization for generalizability and ReLU activation for non-linearity. The final output is ADL feature grids of size *13 x 25 x 6 x 6.*

4.1.2 Pose Backbone Network

The *Pose Backbone Network* extracts scene and scale invariant pose motion action features. The input matches the temporal sampling of the *Video Backbone Network* with 13 frames, each with 25 skeleton joints having x_s,y_s,z_s positions. The *Pose Backbone Network* was designed to consist of parallel paths for nearby, faraway, and positional joint motion features using GCN [167], self-attention [168], and skip connections, respectively, as shown in Figure 10. The parallel branches are concatenated into a single tensor and passed to another GCN stage for joint variant motion features. The reshaped output is 13 frames and 25 channels of *6 x 6* feature grids for multimodal fusion, where each of the 25 channels is associated with a specific human skeleton joint. GCN stages are used to extract motion features independent of the environment by using message passing convolutions between nodes. In this work, human skeletons are transformed to a graph datatype where the nodes represent the 25 discrete skeleton joints and edges represent physical connections between adjacent joints.

For the GCN stage in the parallel section, the input data is the x_s,y_s,z_s position of each skeleton joint which is convolved with positions of adjacent joints for spatial feature extraction. In parallel, the same joint position data is also used in the self-attention module to transfer the data between nodes in the graph using a summation of weights. To determine the weights, the self-attention module assigns each node in the graph a query, key, and value grouping learned during training [168]. For each node, queries are compared to the keys of other nodes and the resulting matching scores are multiplied by the attention node values using a dot-product of weights. The skip connection is a direct data transfer path to pass the positional input data to the next layer. The second GCN stage uses message passing to generate joint dependent motion features. The dimensions of the skeleton joint data after the second GCN stage are *13 x 25 x 36*. Reshaping is then performed on pose motion features in order to match the grid shape of the *Video Backbone Network* output so that the two feature grids can be concatenated in the *Spatial Mid-Fusion Module.*

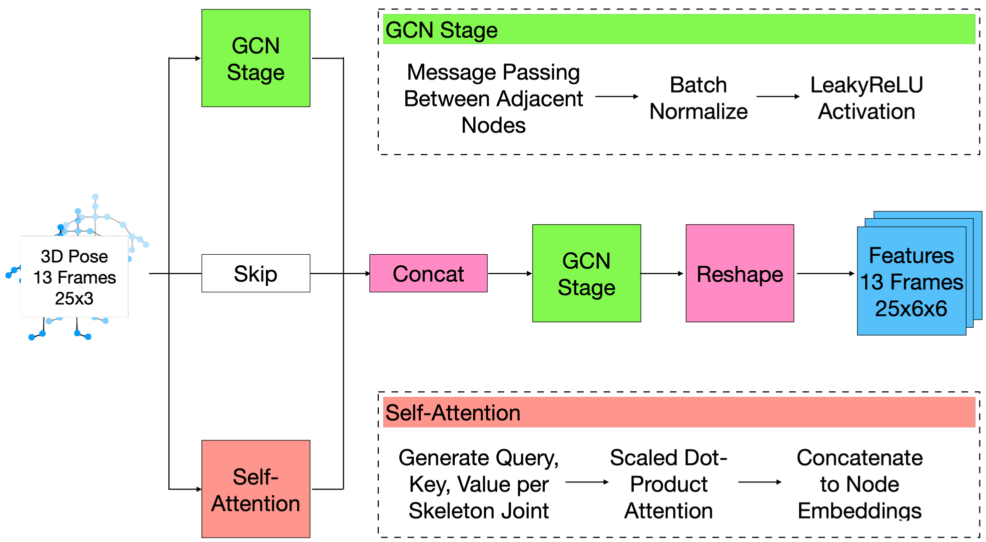


Figure 10 - Pose backbone network architecture [51].

4.1.3 Object Detection Backbone Network

The *Object Detection Backbone* is used to identify and localize objects in the scene during ADL classification. A rolling window approach is used to ensure a new RGB image is acquired with each timestep. The RGB images have an input size of *3 x 1920 x 1080* to use the full resolution available from the video stream to improve detection accuracy. The *Object Detection Network* uses YOLOv5m60 [169] to extract object features from ADL-based home environments, Figure 11. YOLOv5 was selected as it is a state-of-the-art real-time detector for household objects.

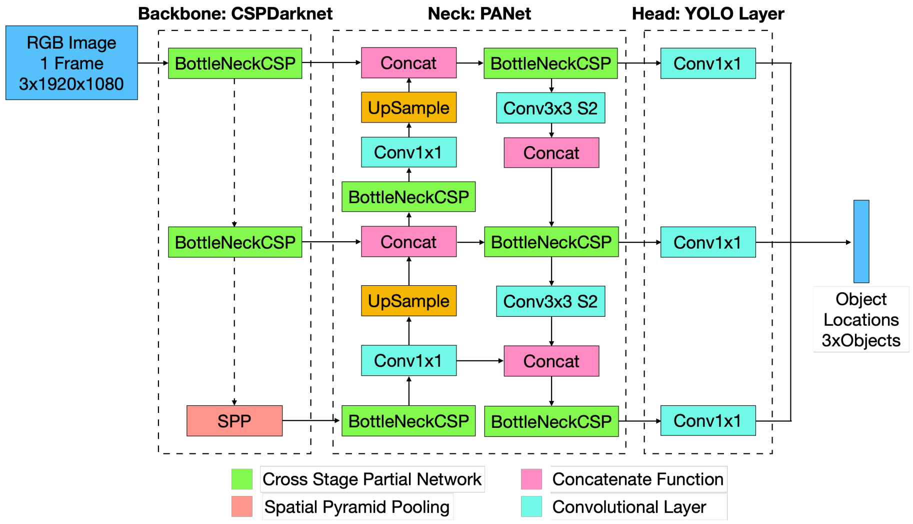


Figure 11 - Object detection backbone network architecture (YOLOv5) [51]. Adapted from [169].

YOLOv5 uses: 1) a Cross Stage Partial (CSP) Network [170] approach to Darknet [171] for extracting high-level spatially invariant features while avoiding unnecessary duplicate gradients, 2) a Path Aggregation Network (PANet) [172] neck layer to use spatial features from each network layer to segment objects, and 3) three individual convolution layers to output object confidence scores. In Darknet, Spatial Pyramid Pooling (SPP) is used to perform information aggregation on inputs with varying sizes [173]. The output of the *Object Detection Backbone Network* is parsed to yield a list of object classes and their x_O, y_O locations. Object classes that were obtained from the COCO dataset [174] included indices 0 (person) and 31-80 (varying household objects, e.g., chair, bottle). A low confidence threshold of 0.25 was used to reduce the potential of false positives from scene and object variation.

4.1.4 Spatial Mid-Fusion Module

A *Spatial Mid-Fusion* module was developed to reshape, scale, and concatenate geometric, motion, and semantic features from the three *Video*, *Pos*e, and *Object Backbone Network* modalities, Figure 12. The size of the input to the *Spatial Mid-Fusion* module is 14 timesteps (13 temporal frames from video/pose and 1 frame for object detection) with 50 channels of feature grids of size *6x6; 14 x 50 x 6 x 6*. The *Spatial Mid-Fusion* consists of: 1) a skip connection for video features to propagate the video feature grids to later layers, and 2) reshaping and scaling on both the skeleton joint motion data for pose features and on the object positions for spatial features. A concatenation step combines video and pose features along the channel dimension and then combines the object location features in the temporal dimension.

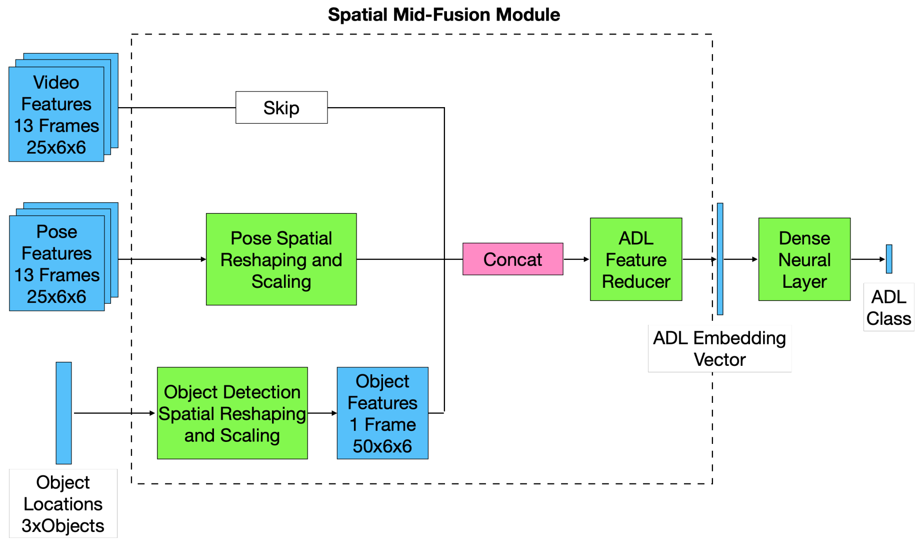


Figure 12 - Spatial mid-fusion, ADL embedding vector, and dense neural layers modules for ADL classification [51].

The pose spatial reshaping and scaling sub-module uses a series of mathematical operations to add spatial context to the pose features relative to the video spatial context, Figure 13. It takes as input the output of the *Pose Backbone Network* (*6 x 6* feature girds *x 25* skeleton joints *x 13* time steps) and uses the x_s,y_s position of each skeleton joint for creating 2D distance maps. For each node, these distance maps are determined by first initializing a Spatial Map *S* of dimensions *6 x 6 x 2* which contains *x,y* positions from -1 to 1 in equal increments. Next, the inverse Euclidian distances are calculated between the normalized x_s,y_s skeleton joint node position and each *x,y* position in *S* to obtain the distance grid *D* of size *6 x 6*. *D* represents the position of the node as a heatmap, where larger values are closer to this node in 2D space. Given a joint feature grid *F*, the new feature grid *F'* is calculated as $F' = D \times F$.

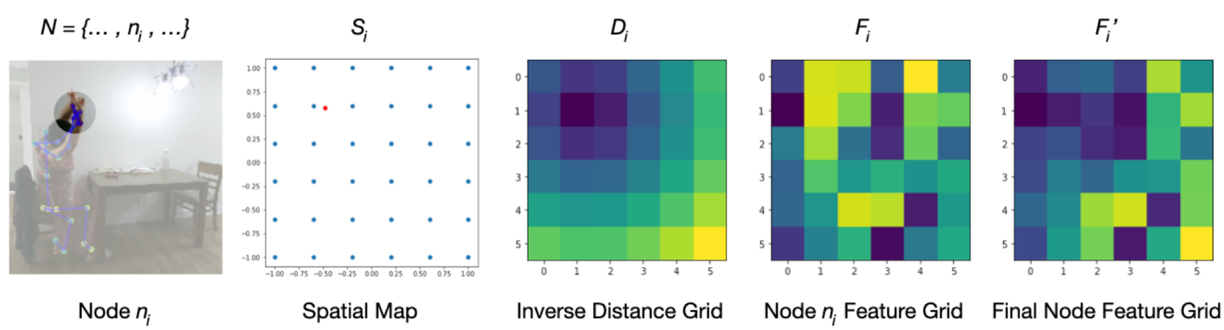


Figure 13 - Pose spatial reshaping and scaling operations.

The object detection spatial reshaping and scaling sub-module uses the list of potential objects and their x_O,y_O locations. For each object, an identity matrix I and a Spatial Map R are initialized identical to S with sizes of 6×6 and $6 \times 6 \times 2$. Next, the inverse Euclidian distances between the object location x_O,y_O and each x,y position in R are used to form the distance grid E of size 6×6. The object feature grid G is then calculated as $G = I \times E$. If multiple objects of the same class exist, object feature grid G' is $G' = G' \times E$, where E is the distance grid for each successive object within the same class.

The new ADL Feature Reducer consists of: 1) a 2D spatial convolution layer for spatial feature extraction between newly fused feature grids, 2) a 1D temporal convolution layer for temporal feature extraction and batch normalization, and 3) (ReLU) activation. Data is then flattened into a single vector of length 25,200 and passed to a linear neural layer to condense the features and create spatio-temporal dependences. Within the linear layer, batch normalization improves generalizability, leaky ReLU [175] activation limits vanishing gradients, and a dropout rate of 0.2 decreases overfitting [176]. The output is the ADL embedding vector of size 128.

ADL Embedding Vector: The ADL Embedding Vector is a low-dimensional representation of a specific ADL containing geometric, motion, and semantic features that are dependent on action timing, locations, motions, and object interactions. The size of the embedding vector follows the dimensionality reduction of the network architecture such that classification accuracy is unaffected. The nature of the embedding space results in ADLs with feature similarity being close in proximity to one another using metrics such as Euclidian distance. The ability to compare features in a low-dimensional space enables contextualization of unseen ADLs based on which existing ADL centroids have the lowest distance to the embedding of the new ADL. Within an ADL class, variations in intra-class embedding vector values determine if an ADL is being performed correctly overtime. ADL embeddings learned from relatively small sets of training data enable generalization within the range of observable features within the dataset. Given that datasets for supervised learning are diverse, the ADL embedding can generalize to new data within the known feature variations, eliminating the need for fully supervised training and decreasing data cost.

4.1.5 Dense Neural Layer

The *Dense Neural Layer* consists of batch normalization and a single fully connected linear layer. These determine scale independent feature interactions within the ADL embedding vector for classification. A dropout rate of 0.5 is used to classify the ADL embedding vector to an ADL class. The output of the *Dense Neural Layer* is the probabilities for each of the ADL classes.

4.1.6 Transfer Learning

Deep transfer learning is used for both the *Video Backbone* and *Object Detection Backbone Networks*. For the *Video Backbone Network*, transfer learning uses the first five layers of X3D small as a spatio-temporal feature extractor. X3D small is pretrained for classification of human activities from the Kinetics dataset [177]. For the *Object Detection Backbone Network*, the entirety of YOLOv5 is pretrained on the COCO dataset [174] for precise location detection of everyday objects in diverse environments.

4.2 Architecture Training

Two variations of the architecture were trained using the ETRI-Activity-3D dataset [178] and the Toyota Smarthome with Refined Skeleton Data V1.2 dataset [179] to show robustness to different datasets. Training used gradient descent based on classification loss.

ETRI-Activity-3D dataset (ETRI): This dataset contains 112,620 samples of 55 activities performed by 50 younger and 50 older adult subjects [178]. Each sample contains an RGB video stream, a depth map, and a skeleton sequence of 3D joint positions. ETRI is chosen as the primary dataset due to its inclusion of activities that directly correspond to typical ADLs performed older adults. It was used for hyperparameter tuning including the depth of layers and the number of convolutional channels. The ADL classes used in training include: eating food with a fork, taking medicine, drinking water, brushing teeth, washing hands, washing face, hanging out laundry, putting on jacket, taking off jacket, putting on/taking off shoes, and brushing hair.

Toyota Smarthome with Refined Skeleton Data V1.2 dataset (Smarthome): This dataset consists of 31 activity classes in 16,000 samples of RGB video, depth video, and human skeleton pose sequences of older adults in smart home environments [179]. All 31 classes were used for training including basic activities such as “take pills” and compounded activities that have a distinct class such as “cook and cleanup” and “cook and cut”. As the pose data from Toyota Smarthome contains 13 skeleton joints rather than 25, architectural modifications to the GCNs were required.

Both datasets were randomized using PyTorch random sampling utilities to ensure an even distribution of ADL classes between the training, validation, and test sets. The data was split into the standard 70% training, 20% validation, and 10% testing sets. Training was accomplished with a learning rate of 2 x 10^{-4}, a batch size of 128 and 20 epochs. Cross entropy [180] was used for classification loss to consider class confidence. The Adam optimizer [181] was used to introduce stochastic behavior for faster convergence using gradient descent. Early stopping was used to select the model with the lowest validation loss. Training loss stabilized after 15 epochs for ETRI and 7 for Smarthome, where training accuracy was 99.9% and 99.7%, respectively. Validation accuracies of 86.9% were obtained for ETRI with optimal hyperparameter selection and 74.1% for Smarthome with more challenging data and without optimized hyperparameters.

4.3 Testing

Several experiments were performed to evaluate the the ADL detection and classification architecture. Network performance is measured by classification accuracy on test sets from the ETRI and Smarthome datasets. The effect of adding individual modalities was determined using an ablation study which compared the multimodal to dual-modal (as primarily used in the literature) and unimodal networks. For evaluating the quality of ADL vector embeddings, the ETRI test set embeddings were used to construct an embedding space for visualization using t-SNE and numerical analysis of distance metrics. Comparison to an embedding space developed solely using RGB video is conducted to measure the impact of multimodality on generating ADL vector embeddings; as embeddings using visual data are a fairly new procedure.

4.3.1 Architecture Testing

To evaluate classification accuracy, the multimodal network was tested on the two large aforementioned ADL datasets. On the ETRI test set with 11 ADL classes (with only basic activities) the low-latency architecture obtained an accuracy of 86.0%, the first to consider real-time applications on ETRI [182]. On the Smarthome test set with high duration variation, basic and compounded activities, and 31 classes, the accuracy obtained was 73.5%. Cross-subject accuracies for Smarthome have been reported to be below 70% [138].

4.3.2 Ablation Study

An ablation study was performed that removed single modalities from the three-modality architecture. Table 5 shows classification accuracy results for ETRI. Multimodality improves model accuracy compared to unimodal and dual-modal networks with the same architecture. The proposed architecture benefits from combining complementary feature data for ADLs to improve classification performance.

Table 5 - Model modality test accuracy [51].

Modality	Test Accuracy
Pose	73.7%
RGB Video	75.1%
Pose and RGB Video	82.4%
Multimodal (Pose+RGB Video+Object)	**86.0%**

4.3.3 ADL Embedding Performance

The ADL embedding vector quality was evaluated using: 1) t-SNE [143] visualization to generate a low-dimensional and high contrast data representation based on neighboring samples by measuring similarities between points in the high-dimensional space, and 2) intra-class variance and inter-class distance. Embedding vectors were created using the ETRI test set and concatenated into the embedding space.

The t-SNE method was used to map the 128 dimensions of the ADL vector embeddings to 2D cartesian plots. ADL vector embeddings from the multimodal network and the RGB video network were compared, Figure 14. RGB video was selected as the unimodal model of comparison since it showed higher accuracy than pose for ETRI as shown in Table 5. The t-SNE visualization shows that the multimodal network has more distinct groupings of similar classes. Namely, when using the RGB video modality, classes with similar environments and large-scale movements such as washing hands or face, and brushing teeth are overlapping in the embedding space (with low separation centroids). Using the multimodal embedding, the centroids have higher separation and visually superior inter-class distinction for similar ADLs. Figure 14 shows distinctions between clothing-based ADLs (putting on/taking off jacket) and consumption-based ADLs (eating, drinking, and taking medicine).

Intra-class variance represents the variance in Euclidian distances of vector embeddings from the same class. On the other hand, inter-class distance measures distances between centroids of classes. Euclidian distance between embeddings was used as the metric as it provides equal weighting of features and computational efficiency [183]. Table 6 shows both intra-class variance and inter-class distance for the multimodal and RGB video embedding spaces. The multimodal embedding space has less variation within classes and greater separation between classes. The lower maximum intra-class variance shows *greater within class* grouping in the embedding space for classes with high levels of activity variability such as drinking water which can occur in many different environments. The higher minimum inter-class distance (by a factor of 1.79) increases separation between the most similar ADL classes within the embedding space.

Contextualization of unseen ADLs was tested by using 5 new samples for each of 5 new ADLs (25 inputs) from the ETRI dataset in the multimodal architecture to obtain their ADL embeddings, Figure 14. These activities included "doing freehand exercises", "spreading bedding/folding bedding", "putting on/taking off glasses", "putting on cosmetics", and "peeling vegetables". The unseen ADLs of "putting on/taking off glasses" and "putting on cosmetics" are near the trained ADL of "brushing hair", as they are similar ADLs with subtle arm movements. However, there is clear distinction between their distributions indicting that they are unique ADLs. For ADLs that have large distributions (e.g., doing freehand exercises), their centroids also show large separations from the centroids of known ADLs, again emphasizing uniqueness.

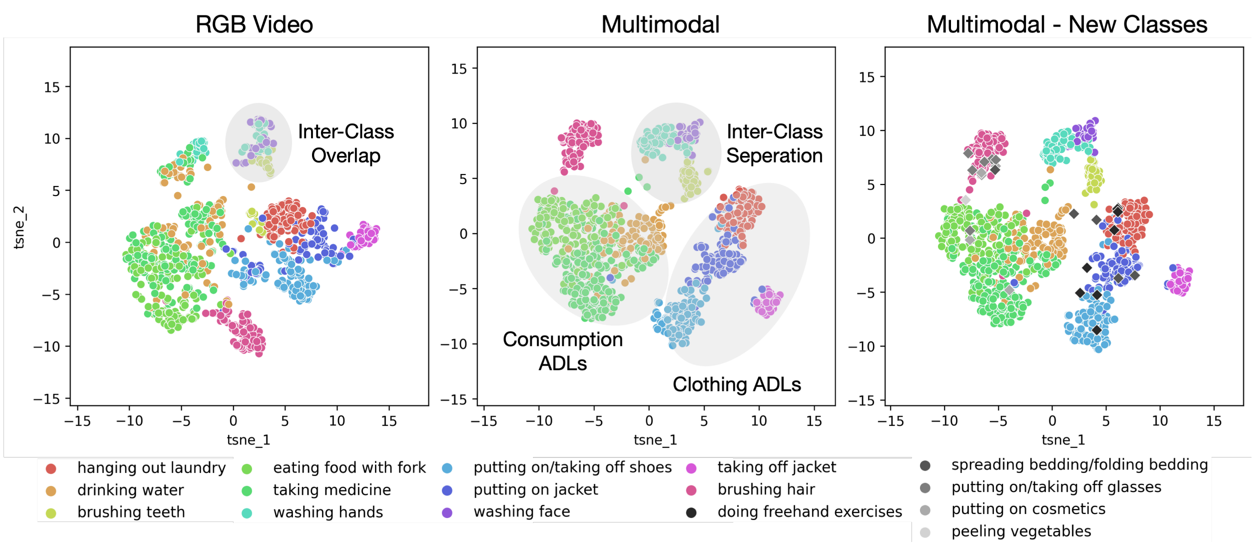


Figure 14 - ADL embedding spaces [51].

Table 6 - Intra-class variation and inter-class distance for embedding spaces [51].

Modality	**Mean Intra-Class Variance**	**Maximum Intra-Class Variance**	**Mean Inter-Class Distance**	**Minimum Inter-Class Distance**
RGB Video	0.67	0.99	3.97	1.90
Multimodal	**0.55**	**0.78**	**4.60**	**3.40**

4.4 Discussion

The objective of ADL classification for SARs is to enable intelligent and autonomous robots capable of perceiving and acting on user behavior. SARs with such capabilities can independently initiate assistive HRI without the prompting of a caregiver or user. Social robots with self-initiation behaviors can promote long-term use [184] and improve ease of use [45] as they interact with users similar to how caregivers do: observing and responding to user actions or needs as they happen. The work herein develops a network that may be used to: 1) classify the ADL task (e.g., eating food with fork, putting on jacket, etc.) in real-time using a light-weight multimodal DL classifier, and 2) classify the ADL type (e.g., seen, unseen, atypically performed) using the ADL embedding space which describes a numeric relationship between an observed ADL and ADLs already in the ADL embedding space (either from training data or seen online). Therefore, using this network, an architecture may be developed that enables a social robot to autonomously recognize and monitor multiple seen, unseen, and atypically performed ADLs in real-time and initiate appropriate assistive HRI.

4.5 Chapter Summary

This chapter presented the first multimodal DL architecture for multi-activity recognition for SARs to enable the proactive initiation of assistive behaviors. The novel architecture introduces the use of an ADL embedding space to uniquely distinguish between a known ADL being performed, a new unseen ADL, or a known ADL being performed atypically in order to assist people in real scenarios. This ADL perception information may be used to initiate robot assistive interactions. An ablation experiment was conducted to show higher ADL classification accuracy for the developed multimodal method over unimodal/dual-modal methods. Visualization of the ADL embedding space shows the inter-class separation necessary for online recognition and monitoring of seen, unseen, and atypically performed ADLs.

CHAPTER 5

5. CONCLUSION AND FUTURE RECOMMENDATIONS

5.1 Summary of Contributions

This thesis aimed to contribute to the field of social robots by developing novel technologies for improved robot capability when assisting users with ADLs. To this end, two main contributions were developed as outlined below.

5.1.1 A Social Robot and Wearable System for Dressing Assistance

The novel socially assistive robot-wearable system was developed to provide dressing assistance using the integration of socially assistive robot and smart clothing technologies. The contributing work developed herein includes a behavior adaptation module to autonomously respond to user dressing states with the socially assistive robot Leia by providing customized assistive behaviors. Experiments conducted on individual modules and the overall robot-wearable system validate its accuracy and robustness in providing reliable dressing assistance. Furthermore, a demonstration-based user study with stakeholders showed positive perceived usefulness, ease of use and attitudes towards the robot-wearable system.

5.1.2 Activity Recognition for SARs

The ADL classification DL architecture for SARs was developed to recognize ADLs and contextualize them with respect other known activities. The novel network: 1) uses RGB-video, 3D user pose locations, and an RGB-image for object detection to extract complementation features and 2) develops an ADL embedding space that provides numerical distance measurements between ADLs. Experiments show the network has improved classification accuracy compared to single or dual modal approaches using an ablation study. Additionally, testing of the ADL embedding space shows it is capable of visually distinguishing new activities not seen during training. Using this ADL recognition on a SAR can enable real-time recognition and monitoring of seen, unseen, and atypically performed ADLs which may be used by the SAR to self-initiate assistive HRI without prior knowledge of which ADL will be performed.

5.2 Recommendations and Future Research

Assistive HRI studies may be conducted with users of diverse cognitive abilities for both of the developed systems to investigate performance with the intended userbase. Robot behaviors for assistive dressing may be improved in terms of user engagement by incorporating dressing performance metrics into robot speech (e.g., motivating the user based on how quickly they have performed a dressing action during previous task completions). Classification accuracy of the activity recognition architecture may be improved by obtaining a larger dataset for training and testing that uses real-world environments. Overall, it is recommended to continue developing new interaction modalities for novel forms of HRI. It is also recommended to develop SAR intelligence architectures that build on ADL task and performance classification to enable long-term monitoring of user ADL ability and/or novel SAR behavior frameworks such as robot empathy.

5.3 Concluding Statement

This thesis presented the development of novel technologies for improving the design of social robots for ADL assistance. The proposed methods included a behavior adaption module to autonomously change robot behaviors based on user preferences during the task of dressing and a deep learning architecture which can enable the first real-time ADL recognition of both seen and unseen ADLs by a robot. Overall, this work explores and develops new ideas for designing social robots that expand the limits of social robot ability and intelligence.

6. REFERENCES

[1] P. F. Edemekong, D. Bomgaars, S. Sukumaran, and S. B. Levy, “Activities of Daily Living,” *StatPearls*, 2019.

[2] C. K. Andersen, K. U. Wittrup-Jensen, A. Lolk, K. Andersen, and P. Kragh-Sørensen, “Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia,” *Health Qual. Life Outcomes*, vol. 2, no. 1, p. 52, 2004, doi: 10.1186/1477-7525-2-52.

[3] D. E. Bloom, D. Canning, and A. Lubet, “Global Population Aging: Facts, Challenges, Solutions & Perspectives,” *Daedalus*, vol. 144, no. 2, pp. 80–92, Apr. 2015, doi: 10.1162/DAED_a_00332.

[4] World Health Organization, “Ageing and health.” WHO Newsroom, Fact Sheets, Oct. 04, 2021. Accessed: Nov. 21, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

[5] “Living in a long-term care home.” Government of Ontario Ministry of Long-Term Care, Dec. 20, 2021. Accessed: Mar. 29, 2022. [Online]. Available: https://www.ontario.ca/page/living-long-term-care-home

[6] K. E. Foley, “Here’s how we can prepare for an aging population.” World Economic Forum. Accessed: Nov. 19, 2021. [Online]. Available: https://www.weforum.org/agenda/2020/02/population-growth-high-demand-caregiving/

[7] N. R. Nicholson, “A Review of Social Isolation: An Important but Underassessed Condition in Older Adults,” *J. Prim. Prev.*, vol. 33, no. 2–3, pp. 137–152, Jun. 2012, doi: 10.1007/s10935-012-0271-2.

[8] S. Brownie, L. Horstmanshof, and R. Garbutt, “Factors that impact residents’ transition and psychological adjustment to long-term aged care: A systematic literature review,” *Int. J. Nurs. Stud.*, vol. 51, no. 12, pp. 1654–1666, Dec. 2014, doi: 10.1016/j.ijnurstu.2014.04.011.

[9] E. Rudnicka, P. Napierała, A. Podfigurna, B. Męczekalski, R. Smolarczyk, and M. Grymowicz, “The World Health Organization (WHO) approach to healthy ageing,” *Maturitas*, vol. 139, pp. 6–11, Sep. 2020, doi: 10.1016/j.maturitas.2020.05.018.

[10] F. Aspinal, J. Glasby, T. Rostgaard, H. Tuntland, and R. G. J. Westendorp, “New horizons: Reablement - supporting older people towards independence,” *Age Ageing*, vol. 45, no. 5, pp. 574–578, Sep. 2016, doi: 10.1093/ageing/afw094.

[11] D.-S. Han, P.-W. Chuang, and E.-C. Chiu, "Effect of home-based reablement program on improving activities of daily living for patients with stroke: A pilot study," *Medicine (Baltimore)*, vol. 99, no. 49, p. e23512, Dec. 2020, doi: 10.1097/MD.0000000000023512.

[12] T. H. Rooijackers *et al.*, "Economic Evaluation of a Reablement Training Program for Homecare Staff Targeting Sedentary Behavior in Community-Dwelling Older Adults Compared to Usual Care: A Cluster Randomized Controlled Trial," *Clin. Interv. Aging*, vol. Volume 16, pp. 2095–2109, Dec. 2021, doi: 10.2147/CIA.S341221.

[13] C. Pettersson and S. Iwarsson, "Evidence-based interventions involving occupational therapists are needed in re-ablement for older community-living people: A systematic review," *Br. J. Occup. Ther.*, vol. 80, no. 5, pp. 273–285, May 2017, doi: 10.1177/0308022617691537.

[14] K. M. Hjelle, H. Tuntland, O. Førland, and H. Alvsvåg, "Driving forces for home-based reablement; a qualitative study of older adults' experiences," *Health Soc. Care Community*, vol. 25, no. 5, pp. 1581–1589, Sep. 2017, doi: 10.1111/hsc.12324.

[15] K. Vik and A. Eide, "Older adults who receive home-based services, on the verge of passivity: the perspective of service providers," *Int. J. Older People Nurs.*, vol. 8, no. 2, pp. 123–130, May 2013, doi: 10.1111/j.1748-3743.2011.00305.x.

[16] J. Morato, S. Sanchez-Cuadrado, A. Iglesias, A. Campillo, and C. Fernández-Panadero, "Sustainable Technologies for Older Adults," *Sustainability*, vol. 13, no. 15, p. 8465, Jul. 2021, doi: 10.3390/su13158465.

[17] T. Vandemeulebroucke, K. Dzi, and C. Gastmans, "Older adults' experiences with and perceptions of the use of socially assistive robots in aged care: A systematic review of quantitative evidence," *Arch. Gerontol. Geriatr.*, vol. 95, p. 104399, Jul. 2021, doi: 10.1016/j.archger.2021.104399.

[18] T. L. Mitzner, T. L. Chen, C. C. Kemp, and W. A. Rogers, "Identifying the Potential for Robotics to Assist Older Adults in Different Living Environments," *Int. J. Soc. Robot.*, vol. 6, no. 2, pp. 213–227, Apr. 2014, doi: 10.1007/s12369-013-0218-7.

[19] A. Umbrico, A. Cesta, G. Cortellessa, and A. Orlandini, "A Holistic Approach to Behavior Adaptation for Socially Assistive Robots," *Int. J. Soc. Robot.*, vol. 12, no. 3, pp. 617–637, Jul. 2020, doi: 10.1007/s12369-019-00617-9.

[20] H. S. Sætra, “The foundations of a policy for the use of social robots in care,” *Technol. Soc.*, vol. 63, p. 101383, Nov. 2020, doi: 10.1016/j.techsoc.2020.101383.

[21] M. Chu, Y.-C. Sun, A. Ashraf, S. F. R. Alves, G. Nejat, and H. E. Naguib, “Making Dressing Easier: Smart Clothes to Help With Putting Clothes on Correctly,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy: IEEE, Oct. 2019, pp. 3981–3986. doi: 10.1109/SMC.2019.8913983.

[22] L. Cobo Hurtado, P. F. Viñas, E. Zalama, J. Gómez-García-Bermejo, J. M. Delgado, and B. Vielba García, “Development and Usability Validation of a Social Robot Platform for Physical and Cognitive Stimulation in Elder Care Facilities,” *Healthcare*, vol. 9, no. 8, p. 1067, Aug. 2021, doi: 10.3390/healthcare9081067.

[23] D. McColl, W.-Y. G. Louie, and G. Nejat, “Brian 2.1: A socially assistive robot for the elderly and cognitively impaired,” *IEEE Robot. Autom. Mag.*, vol. 20, no. 1, pp. 74–83, Mar. 2013, doi: 10.1109/MRA.2012.2229939.

[24] J. M. Beer *et al.*, “The domesticated robot: design guidelines for assisting older adults to age in place,” in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI ’12*, Boston, Massachusetts, USA: ACM Press, 2012, p. 335. doi: 10.1145/2157689.2157806.

[25] G. Bardaro, A. Antonini, and E. Motta, “Robots for Elderly Care in the Home: A Landscape Analysis and Co-Design Toolkit,” *Int. J. Soc. Robot.*, Aug. 2021, doi: 10.1007/s12369-021-00816-3.

[26] S. Frennert, H. Eftring, and B. Östlund, “What Older People Expect of Robots: A Mixed Methods Approach,” in *Social Robotics*, G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, and U. Leonards, Eds., in Lecture Notes in Computer Science, vol. 8239. Cham: Springer International Publishing, 2013, pp. 19–29. doi: 10.1007/978-3-319-02675-6_3.

[27] I. Papadopoulos, C. Koulouglioti, R. Lazzarino, and S. Ali, “Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: a systematic review,” *BMJ Open*, vol. 10, no. 1, p. e033096, Jan. 2020, doi: 10.1136/bmjopen-2019-033096.

[28] W.-Y. G. Louie, D. McColl, and G. Nejat, “Acceptance and Attitudes Toward a Human-like Socially Assistive Robot by Older Adults,” *Assist. Technol.*, vol. 26, no. 3, pp. 140–150, Jul. 2014, doi: 10.1080/10400435.2013.869703.

[29] A. Howard and J. Borenstein, "The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity," *Sci. Eng. Ethics*, vol. 24, no. 5, pp. 1521–1536, Oct. 2018, doi: 10.1007/s11948-017-9975-2.

[30] J. Li, "The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents," *Int. J. Hum.-Comput. Stud.*, vol. 77, pp. 23–37, May 2015, doi: 10.1016/j.ijhcs.2015.01.001.

[31] A. Hong, N. Lunscher, T. Hu, Y. Tsuboi, and X. Zhang, "A Multimodal Emotional Human-Robot Interaction Architecture for Social Robots Engaged in Bidirectional Communication," *IEEE Trans. Cybern.*, p. 15.

[32] W.-Y. G. Louie and G. Nejat, "A learning from demonstration system architecture for robots learning social group recreational activities," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, South Korea: IEEE, Oct. 2016, pp. 808–814. doi: 10.1109/IROS.2016.7759144.

[33] N. Gasteiger, M. Hellou, and H. S. Ahn, "Factors for Personalization and Localization to Optimize Human–Robot Interaction: A Literature Review," *Int. J. Soc. Robot.*, Aug. 2021, doi: 10.1007/s12369-021-00811-8.

[34] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 53, no. 5, pp. 517–527, Oct. 2011, doi: 10.1177/0018720811417254.

[35] S. Bedaf, H. Draper, G.-J. Gelderblom, T. Sorell, and L. de Witte, "Can a Service Robot Which Supports Independent Living of Older People Disobey a Command? The Views of Older People, Informal Carers and Professional Caregivers on the Acceptability of Robots," *Int. J. Soc. Robot.*, vol. 8, no. 3, pp. 409–420, Jun. 2016, doi: 10.1007/s12369-016-0336-0.

[36] S. Serholt, S. Ljungblad, and N. Ní Bhroin, "Introduction: special issue—critical robotics research," *AI Soc.*, pp. s00146-021-01224–x, Jun. 2021, doi: 10.1007/s00146-021-01224-x.

[37] T. Vandemeulebroucke, B. Dierckx de Casterlé, L. Welbergen, M. Massart, and C. Gastmans, "The Ethics of Socially Assistive Robots in Aged Care. A Focus Group Study With Older Adults in Flanders, Belgium," *J. Gerontol. Ser. B*, vol. 75, no. 9, pp. 1996–2007, Oct. 2020, doi: 10.1093/geronb/gbz070.

[38] O. Nocentini, L. Fiorini, G. Acerbi, A. Sorrentino, G. Mancioppi, and F. Cavallo, "A Survey of Behavioral Models for Social Robots," *Robotics*, vol. 8, no. 54, p. 35, 2019, doi: 10.3390/robotics8030054.

[39] C. Moro, G. Nejat, and A. Mihailidis, "Learning and Personalizing Socially Assistive Robot Behaviors to Aid with Activities of Daily Living," *ACM Trans. Hum.-Robot Interact.*, vol. 7, no. 2, pp. 1–25, Oct. 2018, doi: 10.1145/3277903.

[40] L. Robinson *et al.*, "Balancing rights and risks: Conflicting perspectives in the management of wandering in dementia," *Health Risk Soc.*, vol. 9, no. 4, pp. 389–406, Dec. 2007, doi: 10.1080/13698570701612774.

[41] K. Kuca, P. Maresova, B. Klimova, M. Valis, and J. Hort, "Alzheimers disease and language impairments: social intervention and medical treatment," *Clin. Interv. Aging*, p. 1401, Aug. 2015, doi: 10.2147/CIA.S89714.

[42] F. Kumfor, J. L. Hazelton, J. A. Rushby, J. R. Hodges, and O. Piguet, "Facial expressiveness and physiological arousal in frontotemporal dementia: Phenotypic clinical profiles and neural correlates," *Cogn. Affect. Behav. Neurosci.*, vol. 19, no. 1, pp. 197–210, Feb. 2019, doi: 10.3758/s13415-018-00658-z.

[43] J. M. Beer, A. D. Fisk, and W. A. Rogers, "Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction," *J. Hum.-Robot Interact.*, vol. 3, no. 2, p. 74, Jun. 2014, doi: 10.5898/JHRI.3.2.Beer.

[44] M. J. Johnson *et al.*, "Task and Design Requirements for an Affordable Mobile Service Robot for Elder Care in an All-Inclusive Care for Elders Assisted-Living Setting," *Int. J. Soc. Robot.*, vol. 12, no. 5, pp. 989–1008, Nov. 2020, doi: 10.1007/s12369-017-0436-5.

[45] R. Bevilacqua, E. Felici, F. Cavallo, G. Amabili, and E. Maranesi, "Designing Acceptable Robots for Assisting Older Adults: A Pilot Study on the Willingness to Interact," *Int. J. Environ. Res. Public. Health*, vol. 18, no. 20, p. 10686, Oct. 2021, doi: 10.3390/ijerph182010686.

[46] P. Bovbel and G. Nejat, "Casper: An Assistive Kitchen Robot to Promote Aging in Place1," *J. Med. Devices*, vol. 8, no. 3, p. 030945, Sep. 2014, doi: 10.1115/1.4027113.

[47] C. Papadopoulos *et al.*, "The CARESSES Randomised Controlled Trial: Exploring the Health-Related Impact of Culturally Competent Artificial Intelligence Embedded Into

Socially Assistive Robots and Tested in Older Adult Care Homes," *Int. J. Soc. Robot.*, Apr. 2021, doi: 10.1007/s12369-021-00781-x.

[48] O. Zafrani and G. Nimrod, "Towards a Holistic Approach to Studying Human–Robot Interaction in Later Life," *The Gerontologist*, vol. 59, no. 1, pp. e26–e36, Jan. 2019, doi: 10.1093/geront/gny077.

[49] F. Robinson and G. Nejat, "An analysis of design recommendations for socially assistive robot helpers for effective human-robot interactions in senior care," *J. Rehabil. Assist. Technol. Eng.*, vol. 9, p. 205566832211013, Jan. 2022, doi: 10.1177/20556683221101389.

[50] F. Robinson, Z. Cen, H. Naguib, and G. Nejat, "An intelligent socially assistive robot-wearable sensors system for personalized user dressing assistance," *Adv. Robot.*, pp. 1–18, Aug. 2023, doi: 10.1080/01691864.2023.2236170.

[51] F. Robinson and G. Nejat, "A Deep Learning Human Activity Recognition Framework for Socially Assistive Robots to Support Reablement of Older Adults," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, London, United Kingdom: IEEE, May 2023, pp. 6160–6167. doi: 10.1109/ICRA48891.2023.10161404.

[52] D. McColl and G. Nejat, "Meal-Time with a Socially Assistive Robot and Older Adults at a Long-term Care Facility," *J. Hum.-Robot Interact.*, vol. 2, no. 1, pp. 152–171, Mar. 2013, doi: 10.5898/JHRI.2.1.McColl.

[53] C. Pou-Prom, S. Raimondo, and F. Rudzicz, "A Conversational Robot for Older Adults with Alzheimer's Disease," *ACM Trans. Hum.-Robot Interact.*, vol. 9, no. 3, pp. 1–25, Jul. 2020, doi: 10.1145/3380785.

[54] J. Hoorn, "Alice does the MANSA." SELEMCA. Accessed: Jan. 04, 2022. [Online]. Available: http://www.crisprepository.nl/project/selemca/prototype/alice-does-the-mansa

[55] A. K. Pandey and R. Gelin, "A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind," *IEEE Robot. Autom. Mag.*, vol. 25, no. 3, pp. 40–48, Sep. 2018, doi: 10.1109/MRA.2018.2833157.

[56] S. Cooper, A. Di Fava, C. Vivas, L. Marchionni, and F. Ferro, "ARI: the Social Assistive Robot and Companion," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Naples, Italy: IEEE, Aug. 2020, pp. 745–751. doi: 10.1109/RO-MAN47096.2020.9223470.

[57] C. McGinn *et al.*, “Meet Stevie: a Socially Assistive Robot Developed Through Application of a ‘Design-Thinking’ Approach,” *J. Intell. Robot. Syst.*, vol. 98, no. 1, pp. 39–58, Apr. 2020, doi: 10.1007/s10846-019-01051-9.

[58] J. Fasola and M. J. Mataric, “Using Socially Assistive Human–Robot Interaction to Motivate Physical Exercise for Older Adults,” *Proc. IEEE*, vol. 100, no. 8, pp. 2512–2526, Aug. 2012, doi: 10.1109/JPROC.2012.2200539.

[59] D. Gouaillier *et al.*, “The NAO humanoid: a combination of performance and affordability,” *ArXiv08073223 Cs*, Sep. 2008, Accessed: Nov. 18, 2021. [Online]. Available: http://arxiv.org/abs/0807.3223

[60] M. A. Salichs, I. P. Encinar, E. Salichs, Á. Castro-González, and M. Malfaz, “Study of Scenarios and Technical Requirements of a Social Assistive Robot for Alzheimer’s Disease Patients and Their Caregivers,” *Int. J. Soc. Robot.*, vol. 8, no. 1, pp. 85–102, Jan. 2016, doi: 10.1007/s12369-015-0319-6.

[61] D. Fischinger *et al.*, “Hobbit, a care robot supporting independent living at home: First prototype and lessons learned,” *Robot. Auton. Syst.*, vol. 75, pp. 60–78, Jan. 2016, doi: 10.1016/j.robot.2014.09.029.

[62] M. E. Pollack *et al.*, “Pearl: A Mobile Robotic Assistant for the Elderly,” *AAAI Tech. Rep.*, p. 7, 2002.

[63] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, “Relating conversational expressiveness to social presence and acceptance of an assistive social robot,” *Virtual Real.*, vol. 14, no. 1, pp. 77–84, Mar. 2010, doi: 10.1007/s10055-009-0142-1.

[64] K. Saaskilahti, R. Kangaskorte, S. Pieska, J. Jauhiainen, and M. Luimula, “Needs and user acceptance of older adults for mobile service robot,” in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, Paris, France: IEEE, Sep. 2012, pp. 559–564. doi: 10.1109/ROMAN.2012.6343810.

[65] H.-M. Gross *et al.*, “Further progress towards a home robot companion for people with mild cognitive impairment,” in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Seoul, Korea (South): IEEE, Oct. 2012, pp. 637–644. doi: 10.1109/ICSMC.2012.6377798.

[66] L. Taylor *et al.*, “Exploring the applicability of the socially assistive robot Stevie in a day center for people with dementia,” in *2021 30th IEEE International Conference on Robot &*

Human Interactive Communication (RO-MAN), Vancouver, BC, Canada: IEEE, Aug. 2021, pp. 957–962. doi: 10.1109/RO-MAN50785.2021.9515423.

[67] H.-M. Gross *et al.*, "Robot companion for domestic health assistance: Implementation, test and case study under everyday conditions in private apartments," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany: IEEE, Sep. 2015, pp. 5992–5999. doi: 10.1109/IROS.2015.7354230.

[68] O. Avioz-Sarig, S. Olatunji, V. Sarne-Fleischmann, and Y. Edan, "Robotic System for Physical Training of Older Adults," *Int. J. Soc. Robot.*, vol. 13, no. 5, pp. 1109–1124, Aug. 2021, doi: 10.1007/s12369-020-00697-y.

[69] J. Li, W.-Y. G. Louie, S. Mohamed, F. Despond, and G. Nejat, "A user-study with Tangy the Bingo facilitating robot and long-term care residents," in *2016 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, Tokyo, Japan: IEEE, Dec. 2016, pp. 109–115. doi: 10.1109/IRIS.2016.8066075.

[70] C. Thompson, S. Mohamed, W.-Y. G. Louie, J. C. He, J. Li, and G. Nejat, "The robot Tangy facilitating Trivia games: A team-based user-study with long-term care residents," in *2017 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, Ottawa, ON: IEEE, Oct. 2017, pp. 173–178. doi: 10.1109/IRIS.2017.8250117.

[71] N. T. Fitter, M. Mohan, K. J. Kuchenbecker, and M. J. Johnson, "Exercising with Baxter: preliminary support for assistive social-physical human-robot interaction," *J. NeuroEngineering Rehabil.*, vol. 17, no. 1, p. 19, Dec. 2020, doi: 10.1186/s12984-020-0642-5.

[72] C. J. Calo, N. Hunt-Bull, L. Lewis, and T. Metzler, "Ethical Implications of Using the Paro Robot with a Focus on Dementia Patient Care," *AAAI Workshop*, vol. 11, no. 12, p. 5, 2011.

[73] J. Hudson, R. Ungar, L. Albright, R. Tkatch, J. Schaeffer, and E. R. Wicker, "Robotic Pet Use Among Community-Dwelling Older Adults," *J. Gerontol. Ser. B*, vol. 75, no. 9, pp. 2018–2028, Oct. 2020, doi: 10.1093/geronb/gbaa119.

[74] F. Carros *et al.*, "Exploring Human-Robot Interaction with the Elderly: Results from a Ten-Week Case Study in a Care Home," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12. doi: 10.1145/3313831.3376402.

[75] S. Cooper, Ó. Villacañas, L. Marchionni, and F. Ferro, "Robot to support older people to live independently," p. 4.

[76] Alan. S. Brown, "Face to Face with Autism," *Mech. Eng.*, vol. 140, no. 02, pp. 35–39, Feb. 2018, doi: 10.1115/1.2018-FEB-2.

[77] M. A. Salichs *et al.*, "Mini: A New Social Robot for the Elderly," *Int. J. Soc. Robot.*, vol. 12, no. 6, pp. 1231–1249, Dec. 2020, doi: 10.1007/s12369-020-00687-0.

[78] S. A. McGlynn, S. Kemple, T. L. Mitzner, C.-H. A. King, and W. A. Rogers, "Understanding the potential of PARO for healthy older adults," *Int. J. Hum.-Comput. Stud.*, vol. 100, pp. 33–47, Apr. 2017, doi: 10.1016/j.ijhcs.2016.12.004.

[79] S. Baisch *et al.*, "Acceptance of Social Robots by Elder People: Does Psychosocial Functioning Matter?," *Int. J. Soc. Robot.*, vol. 9, no. 2, pp. 293–307, Apr. 2017, doi: 10.1007/s12369-016-0392-5.

[80] S. Tellex, N. Gopalan, H. Kress-Gazit, and C. Matuszek, "Robots That Use Language," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 3, no. 1, pp. 25–55, May 2020, doi: 10.1146/annurev-control-101119-071628.

[81] K. Zsiga, A. Tóth, T. Pilissy, O. Péter, Z. Dénes, and G. Fazekas, "Evaluation of a companion robot based on field tests with single older adults in their homes," *Assist. Technol.*, vol. 30, no. 5, pp. 259–266, Oct. 2018, doi: 10.1080/10400435.2017.1322158.

[82] A. Lee and T. Kawahara, "Recent Development of Open-Source Speech Recognition Engine Julius," in *APSIPA ASC 2009*, Oct. 2009, p. 8. doi: http://hdl.handle.net/2115/39653.

[83] A. Lally, "Natural Language Processing With Prolog in the IBM Watson System," *Assoc. Log. Program. ALP Newsl.*, vol. 9, p. 4, 2011.

[84] P. Torre and J. A. Barlow, "Age-related changes in acoustic characteristics of adult speech," *J. Commun. Disord.*, vol. 42, no. 5, pp. 324–333, Sep. 2009, doi: 10.1016/j.jcomdis.2009.03.001.

[85] J. F. Nussbaum, M. Lee Hummert, A. Williams, and J. Harwood, "Communication and Older Adults," *Ann. Int. Commun. Assoc.*, vol. 19, no. 1, pp. 1–48, Jan. 1996, doi: 10.1080/23808985.1996.11678927.

[86] Paul Ekman Group, "Types of Gestures," *Paul Ekman*. https://www.paulekman.com/nonverbal-communication/types-of-gestures/ (accessed Dec. 20, 2021).

[87] J. B. F. V. Erp and A. Toet, “How to Touch Humans: Guidelines for Social Agents and Robots That Can Touch,” in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Geneva, Switzerland: IEEE, Sep. 2013, pp. 780–785. doi: 10.1109/ACII.2013.145.

[88] PARO Robots, “PARO Manual.” Sep. 2015.

[89] J. Abdi, A. Al-Hindawi, T. Ng, and M. P. Vizcaychipi, “Scoping review on the use of socially assistive robot technology in elderly care,” *BMJ Open*, vol. 8, no. 2, p. e018815, Feb. 2018, doi: 10.1136/bmjopen-2017-018815.

[90] E. Bagheri, O. Roesler, H.-L. Cao, and B. Vanderborght, “A Reinforcement Learning Based Cognitive Empathy Framework for Social Robots,” *Int. J. Soc. Robot.*, vol. 13, no. 5, pp. 1079–1093, Aug. 2021, doi: 10.1007/s12369-020-00683-4.

[91] M. Shao, S. F. D. R. Alves, O. Ismail, X. Zhang, G. Nejat, and B. Benhabib, “You Are Doing Great! Only One Rep Left: An Affect-Aware Social Robot for Exercising,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy: IEEE, Oct. 2019, pp. 3811–3817. doi: 10.1109/SMC.2019.8914198.

[92] C. Moro, S. Lin, G. Nejat, and A. Mihailidis, “Social Robots and Seniors: A Comparative Study on the Influence of Dynamic Social Features on Human–Robot Interaction,” *Int. J. Soc. Robot.*, vol. 11, no. 1, pp. 5–24, Jan. 2019, doi: 10.1007/s12369-018-0488-1.

[93] A. Tapus, C. Tapus, and M. J. Mataric, “The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia,” in *2009 IEEE International Conference on Rehabilitation Robotics*, Kyoto, Japan: IEEE, Jun. 2009, pp. 924–929. doi: 10.1109/ICORR.2009.5209501.

[94] E. Torta, J. Oberzaucher, F. Werner, R. H. Cuijpers, and J. F. Juola, “Attitudes Towards Socially Assistive Robots in Intelligent Homes: Results From Laboratory Studies and Field Trials,” *J. Hum.-Robot Interact.*, vol. 1, no. 2, pp. 76–99, Jan. 2013, doi: 10.5898/JHRI.1.2.Torta.

[95] M. Bajones *et al.*, “Results of Field Trials with a Mobile Service Robot for Older Adults in 16 Private Households,” *ACM Trans. Hum.-Robot Interact.*, vol. 9, no. 2, pp. 1–27, Feb. 2020, doi: 10.1145/3368554.

[96] M. E. Pollack, “Intelligent Technology for an Aging Population The Use of AI to Assist Elders with Cognitive Impairment,” *AI Mag.*, vol. 26, no. 2, p. 16, 2005.

[97] W.-Y. G. Louie and G. Nejat, “A Social Robot Learning to Facilitate an Assistive Group-Based Activity from Non-expert Caregivers,” *Int. J. Soc. Robot.*, vol. 12, no. 5, pp. 1159–1176, Nov. 2020, doi: 10.1007/s12369-020-00621-4.

[98] S. Sabanovic, C. C. Bennett, Wan-Ling Chang, and L. Huber, “PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia,” in *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, Seattle, WA: IEEE, Jun. 2013, pp. 1–6. doi: 10.1109/ICORR.2013.6650427.

[99] H. L. Bradwell, K. J. Edwards, R. Winnington, S. Thill, and R. B. Jones, “Companion robots for older people: importance of user-centred design demonstrated through observations and focus groups comparing preferences of older people and roboticists in South West England,” *BMJ Open*, vol. 9, no. 9, p. e032468, Sep. 2019, doi: 10.1136/bmjopen-2019-032468.

[100] L. Pu, W. Moyle, C. Jones, and M. Todorovic, “The Effectiveness of Social Robots for Older Adults: A Systematic Review and Meta-Analysis of Randomized Controlled Studies,” *The Gerontologist*, vol. 59, no. 1, pp. e37–e51, Jan. 2019, doi: 10.1093/geront/gny046.

[101] T. Zhang, D. B. Kaber, B. Zhu, M. Swangnetr, P. Mosaly, and L. Hodge, “Service robot feature design effects on user perceptions and emotional responses,” *Intell. Serv. Robot.*, vol. 3, no. 2, pp. 73–88, Apr. 2010, doi: 10.1007/s11370-010-0060-9.

[102] I. Kostavelis *et al.*, “Understanding of Human Behavior with a Robotic Agent Through Daily Activity Analysis,” *Int. J. Soc. Robot.*, vol. 11, no. 3, pp. 437–462, Jun. 2019, doi: 10.1007/s12369-019-00513-2.

[103] M. Shao, M. Snyder, G. Nejat, and B. Benhabib, “User Affect Elicitation with a Socially Emotional Robot,” *Robotics*, vol. 9, no. 2, p. 44, Jun. 2020, doi: 10.3390/robotics9020044.

[104] G. Palestra and O. Pino, “Detecting emotions during a memory training assisted by a social robot for individuals with Mild Cognitive Impairment (MCI),” *Multimed. Tools Appl.*, vol. 79, no. 47–48, pp. 35829–35844, Dec. 2020, doi: 10.1007/s11042-020-10092-4.

[105] R. Looije, M. A. Neerincx, and F. Cnossen, “Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors,” *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 6, pp. 386–397, Jun. 2010, doi: 10.1016/j.ijhcs.2009.08.007.

[106] M. Shao, M. Pham-Hung, S. F. R. Alves, M. Snyder, B. Benhabib, and G. Nejat, “Long-Term Exercise Assistance in Group & One-on-One Interactions with a Social Robot & Older Adults,” *ACM Trans. Hum.-Robot Interact. Revis. Rev.*.

[107] W. G. Louie, J. Li, T. Vaquero, and G. Nejat, "A Focus Group Study on the Design Considerations and Impressions of a Socially Assistive Robot for Long-Term Care," *IEEE Int. Symp. Robot Hum. Interact. Commun.*, pp. 237–242, 2014.

[108] Y.-H. Wu, C. Fassert, and A.-S. Rigaud, "Designing robots for the elderly: Appearance issue and beyond," *Arch. Gerontol. Geriatr.*, vol. 54, no. 1, pp. 121–126, Jan. 2012, doi: 10.1016/j.archger.2011.02.003.

[109] Y. Iwamura, M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, "Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets?," in *Proceedings of the 6th international conference on Human-robot interaction - HRI '11*, Lausanne, Switzerland: ACM Press, 2011, p. 449. doi: 10.1145/1957656.1957816.

[110] P. Buono, G. Castellano, B. Decarolis, and N. Macchiarulo, "Social Assistive Robots in Elderly Care: Exploring the role of Empathy," *EMPATHY Empower. People Deal. Internet Things Ecosyst.*, p. 8, 2020.

[111] A. Espingardeiro, "Social assistive robots: a roboethics framework for human robotics interaction in elderly care," *Int. J. Eng. Res. Manag. Stud.*, vol. 2, no. 3, p. 28, Mar. 2015.

[112] J. M. Robillard, I. P. Goldman, T. J. Prescott, and F. Michaud, "Addressing the Ethics of Telepresence Applications Through End-User Engagement," *J. Alzheimers Dis.*, vol. 76, no. 2, pp. 457–460, Jul. 2020, doi: 10.3233/JAD-200154.

[113] A. van Maris, N. Zook, P. Caleb-Solly, M. Studley, A. Winfield, and S. Dogramadzi, "Designing Ethical Social Robots—A Longitudinal Field Study With Older Adults," *Front. Robot. AI*, vol. 7, p. 1, Jan. 2020, doi: 10.3389/frobt.2020.00001.

[114] S. Wong, M. H. Bond, and P. M. Rodriguez Mosquera, "The Influence of Cultural Value Orientations On Self-Reported Emotional Expression Across Cultures," *J. Cross-Cult. Psychol.*, vol. 39, no. 2, pp. 224–229, Mar. 2008, doi: 10.1177/0022022107313866.

[115] C. C. Bennett and S. Šabanović, "The effects of culture and context on perceptions of robotic facial expressions," *Interact. Stud. Soc. Behav. Commun. Biol. Artif. Syst.*, vol. 16, no. 2, pp. 272–302, Nov. 2015, doi: 10.1075/is.16.2.11ben.

[116] S. Coghlan, J. Waycott, A. Lazar, and B. Barbosa Neves, "Dignity, Autonomy, and Style of Company: Dimensions Older Adults Consider for Robot Companions," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, pp. 1–25, Apr. 2021, doi: 10.1145/3449178.

[117] H. R. Lee, H. Tan, and S. Sabanovic, “That robot is not for me: Addressing stereotypes of aging in assistive robot design,” in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, New York, NY, USA: IEEE, Aug. 2016, pp. 312–317. doi: 10.1109/ROMAN.2016.7745148.

[118] B. M. P. Cuff, S. J. Brown, L. Taylor, and D. J. Howat, “Empathy: A Review of the Concept,” *Emot. Rev.*, vol. 8, no. 2, pp. 144–153, Apr. 2016, doi: 10.1177/1754073914558466.

[119] A. Hofmeyer and R. Taylor, “Strategies and resources for nurse leaders to use to lead with empathy and prudence so they understand and address sources of anxiety among nurses practising in the era of COVID-19,” *J. Clin. Nurs.*, vol. 30, no. 1–2, pp. 298–305, Jan. 2021, doi: 10.1111/jocn.15520.

[120] J. C. Pruit, “Getting Dressed,” in *Popular Culture as Everyday Life*, Routledge, 2015, pp. 155–164.

[121] D. F. Mahoney, S. LaRose, and E. L. Mahoney, “Family caregivers’ perspectives on dementia-related dressing difficulties at home: The preservation of self model,” *Dementia*, vol. 14, no. 4, pp. 494–512, Jul. 2015, doi: 10.1177/1471301213501821.

[122] L. Christie, R. Bedford, and A. McCluskey, “Task-specific practice of dressing tasks in a hospital setting improved dressing performance post-stroke: A feasibility study: TASK-SPECIFIC PRACTICE OF DRESSING TASKS,” *Aust. Occup. Ther. J.*, vol. 58, no. 5, pp. 364–369, Oct. 2011, doi: 10.1111/j.1440-1630.2011.00945.x.

[123] G. E. Lancioni *et al.*, “Helping Three Persons with Multiple Disabilities Acquire Independent Dressing through Assistive Technology,” *J. Vis. Impair. Blind.*, vol. 101, no. 12, pp. 768–773, Dec. 2007, doi: 10.1177/0145482X0710101207.

[124] P. Schaad, S. Basler, M. Medini, I. Wissler, T. Bürkle, and M. Lehmann, “The «Intelligent Wardrobe»,” *Nurs. Inform.*, p. 5, 2016.

[125] K. N. Goh, Y. Y. Chen, and E. S. Lin, “Developing a smart wardrobe system,” in *2011 IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, USA: IEEE, Jan. 2011, pp. 303–307. doi: 10.1109/CCNC.2011.5766478.

[126] K. Kalimeri, A. Matic, and A. Cappelletti, “RFID: Recognizing failures in dressing activity,” in *Proceedings of the 4th International ICST Conference on Pervasive Computing*

Technologies for Healthcare, Munchen, Germany: IEEE, 2010. doi: 10.4108/ICST.PERVASIVEHEALTH2010.8896.

[127] E. Ruiz, V. Osmani, L. E. Sucar, and O. Mayora, “Detecting dressing failures using temporal–relational visual grammars,” *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 7, pp. 2757–2770, Jul. 2019, doi: 10.1007/s12652-018-0975-0.

[128] W. Burleson, C. Lozano, V. Ravishankar, J. Lee, and D. Mahoney, “An Assistive Technology System that Provides Personalized Dressing Support for People Living with Dementia: Capability Study,” *JMIR Med. Inform.*, vol. 6, no. 2, p. e21, May 2018, doi: 10.2196/medinform.5587.

[129] S. D. Klee, B. Q. Ferreira, R. Silva, J. P. Costeira, F. S. Melo, and M. Veloso, “Personalized Assistance for Dressing Users,” in *Social Robotics*, A. Tapus, E. André, J.-C. Martin, F. Ferland, and M. Ammi, Eds., in Lecture Notes in Computer Science, vol. 9388. Cham: Springer International Publishing, 2015, pp. 359–369. doi: 10.1007/978-3-319-25554-5_36.

[130] G. Chance, A. Jevtic, P. Caleb-Solly, G. Alenya, C. Torras, and S. Dogramadzi, “‘Elbows Out’—Predictive Tracking of Partially Occluded Pose for Robot-Assisted Dressing,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3598–3605, Oct. 2018, doi: 10.1109/LRA.2018.2854926.

[131] F. Zhang, A. Cully, and Y. Demiris, “Probabilistic Real-Time User Posture Tracking for Personalized Robot-Assisted Dressing,” *IEEE Trans. Robot.*, vol. 35, no. 4, pp. 873–888, Aug. 2019, doi: 10.1109/TRO.2019.2904461.

[132] M. Chu, Y.-C. Sun, A. Ashraf, S. F. R. Alves, G. Nejat, and H. E. Naguib, “Making Dressing Easier: Smart Clothes to Help With Putting Clothes on Correctly,” presented at the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019, pp. 3981–3986. doi: 10.1109/SMC.2019.8913983.

[133] L. Woiceshyn, Y. Wang, G. Nejat, and B. Benhabib, “Personalized clothing recommendation by a social robot,” in *IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, Ottawa, ON: IEEE, Oct. 2017, pp. 179–185. doi: 10.1109/IRIS.2017.8250118.

[134] L. Woiceshyn, Y. Wang, G. Nejat, and B. Benhabib, “A Socially Assistive Robot to Help With Getting Dressed,” in *Design of Medical Devices Conference*, Minneapolis, Minnesota,

USA: American Society of Mechanical Engineers, Apr. 2017, p. V001T11A012. doi: 10.1115/DMD2017-3467.

[135] T. Vogelpohl, C. Beck, P. Heacook, and S. Mercer, "'I Can Do It!' Dressing Promoting Independence Through Individualized Strategies," *J. Gerontiological Nurs.*, pp. 39–42, Mar. 1996.

[136] S. Das, S. Sharma, R. Dai, F. Bremond, and M. Thonnat, "VPN: Learning Video-Pose Embedding for Activities of Daily Living." arXiv, Jul. 06, 2020. Accessed: Jul. 12, 2022. [Online]. Available: http://arxiv.org/abs/2007.03056

[137] S. Das, R. Dai, D. Yang, and F. Bremond, "VPN++: Rethinking Video-Pose embeddings for understanding Activities of Daily Living," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2021, doi: 10.1109/TPAMI.2021.3127885.

[138] H. Kim, D. Kim, and J. Kim, "Learning Multi-modal Attentional Consensus in Action Recognition for Elderly-Care Robots," in *2021 18th International Conference on Ubiquitous Robots (UR)*, Gangneung, Korea (South): IEEE, Jul. 2021, pp. 308–313. doi: 10.1109/UR52253.2021.9494666.

[139] A. Ghods and D. J. Cook, "Activity2Vec: Learning ADL Embeddings from Sensor Data with a Sequence-to-Sequence Model." arXiv, Jul. 12, 2019. Accessed: Sep. 09, 2022. [Online]. Available: http://arxiv.org/abs/1907.05597

[140] Y. Jain, C. I. Tang, C. Min, F. Kawsar, and A. Mathur, "ColloSSL: Collaborative Self-Supervised Learning for Human Activity Recognition," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 1, pp. 1–28, Mar. 2022, doi: 10.1145/3517246.

[141] M. T. H. Tonmoy, S. Mahmud, A. K. M. M. Rahman, M. A. Amin, and A. A. Ali, "Hierarchical Self Attention Based Autoencoder for Open-Set Human Activity Recognition." arXiv, Mar. 07, 2021. Accessed: Jul. 12, 2022. [Online]. Available: http://arxiv.org/abs/2103.04279

[142] J. Sung, I. Lenz, and A. Saxena, "Deep multimodal embedding: Manipulating novel objects with point-clouds, language and trajectories," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, Singapore: IEEE, May 2017, pp. 2794–2801. doi: 10.1109/ICRA.2017.7989325.

[143] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[144] H. Mahdi, S. A. Akgun, S. Saleh, and K. Dautenhahn, "A survey on the design and evolution of social robots — Past, present and future," *Robot. Auton. Syst.*, vol. 156, p. 104193, Oct. 2022, doi: 10.1016/j.robot.2022.104193.

[145] S. F. R. Alves, M. Shao, and G. Nejat, "A Socially Assistive Robot to Facilitate and Assess Exercise Goals," in *Proceedings of the IEEE International Conference of Robotics and Automation Workshop on Mobile Robot Assistants for the Elderly*, Montreal, QC, Canada, 2019, p. 5.

[146] N. Villaroman, D. Rowe, and B. Swan, "Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor," in *Proceedings of the 2011 conference on Information technology education - SIGITE '11*, West Point, New York, USA: ACM Press, 2011, p. 227. doi: 10.1145/2047594.2047654.

[147] T. G. Dietterich, "Hierarchical Reinforcement Learning with the MAXQ Value Function Decomposition," *arXiv:cs/9905014*, May 1999, [Online]. Available: http://arxiv.org/abs/cs/9905014

[148] Paul Ekman Group, "Types of Gestures," *Paul Ekman*. https://www.paulekman.com/nonverbal-communication/types-of-gestures/ (accessed Dec. 20, 2021).

[149] K. Kellermann and T. Cole, "Classifying Compliance Gaining Messages: Taxonomic Disorder and Strategic Confusion," *Commun. Theory*, vol. 4, no. 1, pp. 3–60, Feb. 1994, doi: 10.1111/j.1468-2885.1994.tb00081.x.

[150] S. Saunderson and G. Nejat, "It Would Make Me Happy if You Used My Guess: Comparing Robot Persuasive Strategies in Social Human–Robot Interaction," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1707–1714, Apr. 2019, doi: 10.1109/LRA.2019.2897143.

[151] "How to support a person with dementia to get dressed or change clothes," *Alzheimer's Society*, Mar. 08, 2021. https://www.alzheimers.org.uk/get-support/daily-living/getting-dressed-changing-clothes

[152] Q. Lin *et al.*, "E-Jacket: Posture Detection with Loose-Fitting Garment using a Novel Strain Sensor," in *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, Apr. 2020, pp. 49–60. doi: 10.1109/IPSN48710.2020.00-47.

[153] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, “Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model,” *Int. J. Soc. Robot.*, vol. 2, no. 4, pp. 361–375, Dec. 2010, doi: 10.1007/s12369-010-0068-5.

[154] I. H. Kuo *et al.*, “Age and gender factors in user acceptance of healthcare robots,” in *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan: IEEE, Sep. 2009, pp. 214–219. doi: 10.1109/ROMAN.2009.5326292.

[155] C. Getson and G. Nejat, “The Robot Screener Will See You Now: A Socially Assistive Robot for COVID-19 Screening in Long-Term Care Homes,” in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Napoli, Italy: IEEE, Aug. 2022, pp. 672–677. doi: 10.1109/RO-MAN53752.2022.9900620.

[156] B. Khayamian Esfahani, “THE IMPORTANCE OF GENDER-AWARE DESIGN IN DIGITAL HEALTH WEARABLES: A CO-DESIGN STUDY FOSTERING SUN PROTECTION BEHAVIOUR IN YOUNG MEN,” *Proc. Des. Soc.*, vol. 1, pp. 3031–3040, Aug. 2021, doi: 10.1017/pds.2021.564.

[157] M. Dehghani, K. J. Kim, and R. M. Dangelico, “Will smartwatches last? factors contributing to intention to keep using smart wearable technology,” *Telemat. Inform.*, vol. 35, no. 2, pp. 480–490, May 2018, doi: 10.1016/j.tele.2018.01.007.

[158] W.-Y. G. Louie, D. McColl, and G. Nejat, “Acceptance and Attitudes Toward a Human-like Socially Assistive Robot by Older Adults,” *Assist. Technol.*, vol. 26, no. 3, pp. 140–150, Jul. 2014, doi: 10.1080/10400435.2013.869703.

[159] M. Pino, M. Boulay, F. Jouen, and A.-S. Rigaud, “‘Are we ready for robots that care for us?’ Attitudes and opinions of older adults toward socially assistive robots,” *Front. Aging Neurosci.*, vol. 7, Jul. 2015, doi: 10.3389/fnagi.2015.00141.

[160] K. Winkle, P. Caleb-Solly, A. Turton, and P. Bremner, “Mutual Shaping in the Design of Socially Assistive Robots: A Case Study on Social Robots for Therapy,” *Int. J. Soc. Robot.*, vol. 12, no. 4, pp. 847–866, Aug. 2020, doi: 10.1007/s12369-019-00536-9.

[161] K. D. Elsbach and I. Stigliani, “New Information Technology and Implicit Bias,” *Acad. Manag. Perspect.*, vol. 33, no. 2, pp. 185–206, May 2019, doi: 10.5465/amp.2017.0079.

[162] R. E. McGrath, M. Mitchell, B. H. Kim, and L. Hough, “Evidence for response bias as a source of error variance in applied assessment.,” *Psychol. Bull.*, vol. 136, no. 3, pp. 450–470, May 2010, doi: 10.1037/a0019216.

[163] C. Feichtenhofer, “X3D: Expanding Architectures for Efficient Video Recognition.” arXiv, Apr. 09, 2020. Accessed: Jul. 12, 2022. [Online]. Available: http://arxiv.org/abs/2004.04730

[164] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition.” arXiv, Dec. 10, 2015. Accessed: Aug. 19, 2022. [Online]. Available: http://arxiv.org/abs/1512.03385

[165] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.” arXiv, Mar. 02, 2015. Accessed: Aug. 19, 2022. [Online]. Available: http://arxiv.org/abs/1502.03167

[166] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU).” arXiv, Feb. 07, 2019. Accessed: Sep. 09, 2022. [Online]. Available: http://arxiv.org/abs/1803.08375

[167] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks.” arXiv, Feb. 22, 2017. Accessed: Aug. 18, 2022. [Online]. Available: http://arxiv.org/abs/1609.02907

[168] A. Vaswani *et al.*, “Attention Is All You Need.” arXiv, Dec. 05, 2017. Accessed: Aug. 04, 2022. [Online]. Available: http://arxiv.org/abs/1706.03762

[169] G. Jocher, “ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference.” Zenodo, Feb. 22, 2022. [Online]. Available: doi:10.5281/zenodo.6222936

[170] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, “CSPNet: A New Backbone that can Enhance Learning Capability of CNN.” arXiv, Nov. 26, 2019. Accessed: Aug. 18, 2022. [Online]. Available: http://arxiv.org/abs/1911.11929

[171] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger.” arXiv, Dec. 25, 2016. Accessed: Sep. 15, 2022. [Online]. Available: http://arxiv.org/abs/1612.08242

[172] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path Aggregation Network for Instance Segmentation.” arXiv, Sep. 18, 2018. Accessed: Aug. 18, 2022. [Online]. Available: http://arxiv.org/abs/1803.01534

[173] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” 2014, pp. 346–361. doi: 10.1007/978-3-319-10578-9_23.

[174] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., in Lecture Notes in Computer Science, vol. 8693. Cham: Springer International Publishing, 2014, pp. 740–755. doi: 10.1007/978-3-319-10602-1_48.

[175] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical Evaluation of Rectified Activations in Convolutional Network.” arXiv, Nov. 27, 2015. Accessed: Sep. 12, 2022. [Online]. Available: http://arxiv.org/abs/1505.00853

[176] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15.1, pp. 1929–1958, 2014.

[177] J. Carreira and A. Zisserman, “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 4724–4733. doi: 10.1109/CVPR.2017.502.

[178] J. Jang, D. Kim, C. Park, M. Jang, J. Lee, and J. Kim, “ETRI-Activity3D: A Large-Scale RGB-D Dataset for Robots to Recognize Daily Activities of the Elderly,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 10990–10997. doi: 10.1109/IROS45743.2020.9341160.

[179] S. Das *et al.*, “Toyota Smarthome: Real-World Activities of Daily Living,” *IEEE Int. Conf. Comput. Vis. ICCV*, p. 10, Oct. 2019.

[180] Z. Zhang and M. Sabuncu, “Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels,” *Adv. Neural Inf. Process. Syst.*, vol. 31, p. 11, 2018.

[181] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization.” arXiv, Jan. 29, 2017. Accessed: Sep. 09, 2022. [Online]. Available: http://arxiv.org/abs/1412.6980

[182] X. Shu, J. Yang, R. Yan, and Y. Song, “Expansion-Squeeze-Excitation Fusion Network for Elderly Activity Recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 8, pp. 5281–5292, Aug. 2022, doi: 10.1109/TCSVT.2022.3142771.

[183] P.-E. Danielsson, “Euclidean distance mapping,” *Comput. Graph. Image Process.*, vol. 14, no. 3, pp. 227–248, Nov. 1980, doi: 10.1016/0146-664X(80)90054-4.

[184] C. Clabaugh and M. Matarić, “Escaping Oz: Autonomy in Socially Assistive Robotics,” *Annu. Rev. Control Robot. Auton. Syst.*, vol. 2, no. 1, pp. 33–61, May 2019, doi: 10.1146/annurev-control-060117-104911.