Dementia Scale Classification Based on Ubiquitous Daily Activity and Interaction Sensing

Shogo Okada
Japan Advanced Institute of Science and Technology,
RIKEN AIP, Japan
okada-s@jaist.ac.jp

Ken Inoue
George and Shaun, Co. Ltd, Japan
ken.inoue@george-shaun.com

Toru Imai, Mami Noguchi
NIPPON TELEGRAPH AND TELEPHONE WEST CORPORATION, Japan
{mami.noguchi.ph, tooru.imai.tv}@west.ntt.co.jp

Kaiko Kuwamura
Sharp Corporation, Japan
kuwamura.kaiko@sharp.co.jp

Abstract—**This paper investigates the integration of different approaches to automatically predict high/low-score on the dementia scale. We propose two different approaches to predict this value by capturing the following: (1) the participant's interaction behavior with a humanoid robot and (2) the indoor daily activity in the residence using ubiquitous sensors. The interaction and indoor activity data set were obtained by recording 32 participants living in common residences, including 17 with symptoms of dementia, as indicated through a cognitive test (Revised Hasegawa Dementia Scale). To obtain the interaction features, we extracted the turn-taking features of interaction with a mobile-typed humanoid robot. To extract the indoor activity features, we collected the location data of each participant in the residence using the received signal strength indicators (RSSIs) of Bluetooth signals from different access points (e.g., shared spaces or the participant's room). In the experimental evaluation, we trained binary classification models for classifying the score on the dementia scale from these datasets. The results show that the best classification accuracy (0.875) is achieved when interaction and activity features are fused using a random forest classifier.**

Index Terms—**Dementia, Ubiquitous indoor positioning, Human robot interaction, Machine learning**

I. Introduction

Dementia is a syndrome comprising brain diseases that gradually decrease in the patient's ability to think and remember to the level of affecting their daily functioning [1]. Dementia is also known to have a strong emotional component and the detection of indicators of dementia clearly help us understand the emotional component. In recent years, many works have focused on the automatic detection of dementia based on a patient's verbal communication abilities, such as speech, language attributes, and interaction with computer avatars [2], [3], along with features of physical activity such as walking speed [4]. Many studies have focused on the detection of dementia through sensing either interaction behaviors or physical activity. The effectiveness of the integration of both types of information has been unexplored.

In this research, we conducted a comparative analysis of automatically predicting scores on the dementia scale using multiactivity features based on the participant's interaction behaviors, which was observed through human-robot interaction under a noncontrolled condition setting (interaction dataset), and their indoor daily activity, which was observed using an indoor positioning system in the participants'residences (daily activity dataset). To observe and record their daily activity, an indoor positioning system was used, and each of the participants were equipped with a mobile beacon. We collected the location data of the participants in their residences using the received signal strength indicators (RSSIs) of Bluetooth signals from different access points.

The 32 participants and their families agreed to the recording of the dataset. All 32 participants completed a cognitive test called the Revised Hasegawa's Dementia Scale (HDS-R) [5], which has been proposed to screen for dementia. We extracted the turn-taking features from the interaction data, including the reaction time after the questions and the speaking length. The location data were converted into indoor activity features, capturing how much time each participant stayed in each room or shared space. For this experiment, classification models are trained using the obtained dataset that would allow us to detect the possibility of dementia as a binary classification task by distinguishing between a higher-scale group and a lower-scale group. The main contributions of this study are summarized below.

Multiactivity data set for analyzing the dementia scale: We collected a novel multiactivity dataset from the participants in residential facilities over three months to extract interaction behaviors and indoor daily activities. Long-term data collection under these realistic conditions enabled us to model the behaviors of the participants.

Fusing interaction and indoor activity for detecting dementia: This study addresses a novel challenge in investigating the possibility of automatic prediction of the dementia scale by integrating location-based activity analysis using ubiquitous sensing and interaction behavior analysis.

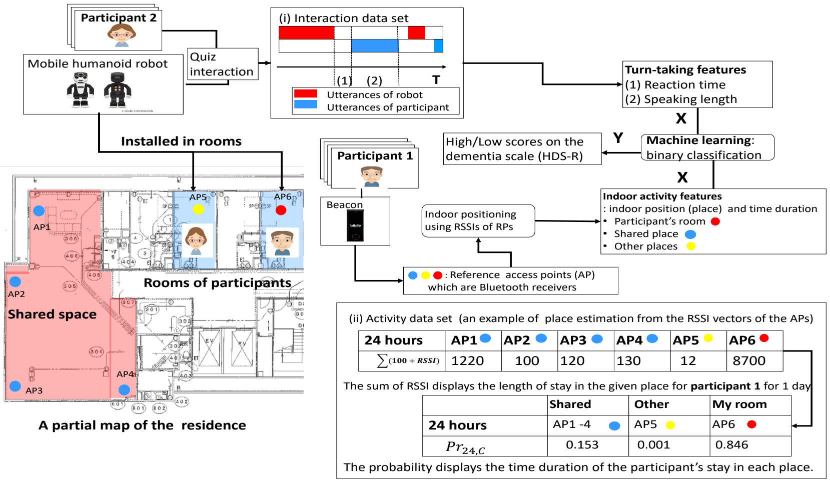
Automatic dementia scale classification: Collecting the interaction and location datasets, extracting the features from these datasets was conducted in a fully automatic manner.

II. Related works

A. Analyzing Dementia using Interaction Features

Orimaye et al. [6] reported the effectiveness of using linguistic features including syntactic features to identify people with Alzheimer's disease. Boschi et al. [7] reviewed the

978-1-7281-3888-6/19/$31.00 ©2019 IEEE



(ii) Activity data set (an example of place estimation from the RSSI vectors of the APs)

24 hours	AP1 ●	AP2 ●	AP3 ●	AP4 ●	AP5 ●	AP6 ●
$\sum(100 + RSSI)$	1220	100	120	130	12	8700

The sum of RSSI displays the length of stay in the given place for **participant 1** for 1 day

	Shared	Other	My room
24 hours	AP1 -4 ●	AP5 ●	AP6 ●
$Pr_{24,C}$	0.153	0.001	0.846

The probability displays the time duration of the participant's stay in each place.

Fig. 1. Overview of dementia scale score prediction scheme

TABLE I
STATISTICS OF PARTICIPANTS AND THE NUMBER OF PARTICIPANT'S IN HIGH/LOW-SCORE GROUP OF DEMENTIA SCALE

	Num. of Participants	HDS-R [High/Low]
Interaction	19	[12/7]
Location	19	[7/12]
Both	6	[4/2]
Total	32	[15/17]
Gender [F/M]	[28/4]	
Average age	84.56 (±5.25)	

differences in the tasks used to elicit connected speech, picture description, story narration, and interviews, considering the possible different contributions to the assessment of different linguistic domains. Masrani et al. [8] investigated the forms of blog posts by writers with dementia using natural language processing.

König et al. [9] proposed an approach to detecting dementia from vocal features using four cognitive vocal tasks. In addition, König et al. [10] collected automatic speech analytics using a mobile application and showed its effectiveness for use in the assessment task. Fraser et al. [11] proposed an approach to identifying Alzheimer's disease through speech recordings from the DementiaBank dataset by using multimodal features including linguistic and acoustic features from the transcript. Aramaki et al. [12] analyzed written and spoken narratives to compare the language abilities of study participants with and without mild cognitive impairment (MCI) to explore the relationship between cognitive and language abilities. Regarding visual features, some studies have reported mutual gazes [13] and facial expressions [14], [15] as key descriptors for use in detecting dementia.

In recent years, agent systems with multimodal or social signal sensing have been developed for counseling participants [16], [3]. This research is inspired from the findings of [3]. We extracted the reaction time (the gap feature) after a question from a robot and the speaking length, and we used these features to detect dementia. Furthermore, previous studies focused on observing the communication behaviors to detect dementia. Thus, the participants'daily activity was not considered. Herein, we investigate the relationships between interaction behavior and dementia and between daily activity and dementia.

B. Detecting physiological disease from physical activity

Hodges et al. [17] used wearable devices (RFID bracelets) and RFID- tagged objects to detect indications of cognitive impairments such as dementia and traumatic brain injury by monitoring individuals performing a well-defined routine task—making coffee.

Hayes et al. [4] investigated the association between walking speed, the amount of daily activity in the participants'residences and the level (high/low) of MCI by extracting activity features using a motion sensor system. Dawadi et al. [18] investigated the association between the ability of a participant to complete an activity and the health assessment (dementia or cognitively healthy). Riboni et al. [19] investigated the ability of individuals to perform activities independently without assistance as an important feature in estimating their functional health. Robben et al. [20] developed an ambient sensor monitoring system and collected sensor data in a participant's residence over three years. In addition, [21] proposed an algorithm for quantifying the changes in everyday behavior and evaluated the algorithm using a longitudinal sensor dataset.

In many previous studies, datasets were collected using a smart-home test bed or in a laboratory setting. We focus on extracting the participants'activity during daily life (ADL), using only location sensors in the participants'residences. This means that the data collection process is done in a real-life environment. The main difference between this study and previous research is that our dataset includes both data on the interaction with the robot.

III. METHOD

A. Data recording in the nursing residential facility

Figure 1 shows the overview of the dementia scale score prediction. We recorded the interaction and indoor daily activity datasets in two nursing residential facilities in Japan. We recruited 32 Japanese participants. The Research Ethics Committees of "NIPPON TELEGRAPH AND TELEPHONE WEST CORPORATION" reviewed and approved the collection of data and the corresponding research using this dataset. The dataset, excluding personal information (age, gender, name, and audio) that could be used by a third party to identify and discriminate against the participants, were shared only to all coauthors of this study[1]. The statistics of age and gender were shared only in that the average age of the participants

[1] This dataset is not publicly available.

was 84.56 (±5.25) and there were 4 male and 28 female participants. Written informed consent was obtained from all participants or from a capable family member before collecting the following data.

The participants initially completed the dementia screening test called the Revised Hasegawa's Dementia Scale (HDSR) test (Section III-B) in the residential facility. Eleven participants out of the 32 were unable to complete the test. An expert caregiver analyzed the test data for these 11 participants and examined the participants'cognitive health. Their scores were estimated as below 20, and their cognitive function was judged as decreased. In the HDS-R, if the score is below 20 points (cut-off), the possibility of dementia is high. This study investigates the association between multiple activities and the score on the HDS-R. Table I shows a summary of the statistics of the participants in the high/low-score groups of the dementia scale.

In total, 19 participants agreed to install the mobile robot into their room and to record the interaction log data, including the timing of the participants'speaking turns calculated via a voice-activity detection technique. A total of 1056 sessions of interaction was recorded over three months. Each session included 1-10 turns taken between the robot and the participant, and the average was 52.8 sessions per participant. A total of 19 participants agreed to always carry a mobile beacon and to record the RSSI signal sent out from the beacon. Indoor location activity data were recorded for all participants over 854 days, and average was 44.9 days per participant. In total, 6 participants agreed to record both kinds of data.

B. Revised Hasegawa's Dementia Scale (HDS-R)

The Revised Hasegawa's dementia scale (HDS-R) consists of nine simple questions (including questions to check for memory and simple mathematical-logical capacity), with a maximum score of 30. The cutoff point for the age-associated dementia screening is 20-21. A lower score means high risk of dementia. The test's effectiveness in screening for age-associated dementia was examined [5].

Some questions in HDS-R are common to that in Mini-Mental Status Examination (MMSE) which is most frequently used in worldwide. The HDS-R correlation coefficient relative to the MMSE was as high as 0.94. Kim et al. [22] reported that the diagnostic accuracy of HDS-R was significantly higher than that of the MMSE, regardless of the educational level of the subjects, as a result of their comparison between the diagnostic accuracies for Alzheimer's disease. From this findings, HDS-R is often used for Japanese people and we used the HDS-R as the dementia scale in this study.

C. Interaction dataset

To collect the interaction data, we used the mobile robot RoBoHon (SR-01M-W) [2], which was produced by Sharp Corporation. The robot is a humanoid-type small robot (19.5 cm and 390 g) and has the functions of a mobile smart-phone. The specs are as follows: OS - Android 5.0 and CPU - Qualcomm Snapdragon 400 processor at 1.2 GHz. Though it also has a camera and microphone similar to those of a smart-phone, the camera was not used in this study, and audio data was used only for voice activity detection. For this experiment, a dialog system was developed in the robot using a software development kit for speech processing and dialog management. The objective within the dialog system is to play a quiz game with the participant. The system has two dialog modes: (1) small-talk including greeting and self-disclosure; and (2) quiz, including questions in a cognitive test. The dialog procedure is shown below.

[Step 1]: The system starts to talk to the participant (e.g., The system says "Hello (the participant's name)!"). If a voiced reaction is detected via voice-activity detection (VAD) within TGs after the system utterance, then the system goes to Step 2. TG is the maximum time for waiting for the utterance.

[Step 2]: The system continues small-talk (e.g., "How are you? ", "Talk with me") without speech recognition of the participant's utterance. The system automatically goes to Step 3 after this step.

[Step 3]: The system starts the quiz (e.g. Q1: "What year is it?"). If the voiced reaction is recognized within TGs after the system utterance, then repeat Step 3 with the next question. Otherwise, go to Step 4.

[Step 4]: The system continues small-talk (e.g., "Hi (the participant's name), are you there? "). If the voiced reaction is recognized within TGs after system utterance, the system goes back to Step 3. Otherwise, the application is terminated after a greeting from the system.

We set TG to 20 seconds. The robot attempts to talk with the participant every two hours, and each session may include up to two quizzes (Step 1-4). A total of twenty quizzes was prepared for the dialog, and two quizzes were chosen randomly for each session. If a quiz is used for a session, the same quiz is not used in the next session in the day. We recorded the timestamps (start and end time) of the utterances of the system and the participants via VAD.

D. Indoor location dataset

Although the two nursing residential facilities hosting the participants are very different, they both have individual rooms with a bed and a bath room and a shared space for free communication between the residents. To collect the indoor location data, we used a Bluetooth low-energy (BLE) beacon, Biblle [3], which was produced by George and Shaun, Co. Ltd. This Bluetooth beacon has a small radio transmitter (6 × 0.6 × 2.2 cm, 9.07 g) that sends out signals within a radius of 10-30 meters (interior spaces). These beacons are cost-effective, can be installed with minimal effort. We installed the reference access points (AP), which are receivers for the Bluetooth signal, in the participants'rooms and the shared spaces of the residential facilities. We estimate the position of a participant with a beacon using the RSSI of the Bluetooth signal and the location coordinates of the APs.

[2] https://robohon.com/global/index.php

[3] https://george-shaun.com/

IV. Multi-activity feature extraction

For extracting the interaction features of i th participant, the feature set is extracted per session from the total $\mathbf{S_i}$ sessions in which the participant talked with the robot. One session is defined as the time from the beginning of the dialog with the robot to the end of the dialog. For the indoor activity features, the feature set is extracted per day out of $\mathbf{D_i}$ days that the location data was observed from the participant.

A. Interaction activity with the robot

We extracted the turn-taking features, which are effective for use in detecting dementia. Only turn-taking in the quiz was used for feature extraction in this study. If voice activity from the participant was detected in the quiz, the robot-participant utterance pair was defined as a turn taken. Webrtc open source API [4] was used for voice activity detection in this study.

1) When a participant's (t) th utterance in the (i) th session $U_{p,t,i}$ is detected via VAD, let the end time of the system utterance of the quiz be $ET_{r,t,i}$ and the start, and end times of $U_{p,t,i}$ are $ST_{p,t,i}$ and $ET_{p,t,i}$.
2) The reaction time duration is calculated as $RT_{t,i} = ST_{p,t,i} - ET_{r,t,i}$.
3) The speaking length within the turn is calculated as $SL_{t,i} = ET_{p,t,i} - ST_{p,t,i}$.

The turn-taking features are calculated as follows.

Reaction time: First, the statistical parameters including the mean, standard deviation, maximum, minimum, median, percentile:75% (p_{75}), percentile:25% (p_{25}) and difference between percentiles $(p_{75} - p_{25})$ of the reaction time $RT_{t,i}$ is calculated over all utterance. Second, the statistical parameters including the mean $(Ses.Mean_{RT})$, standard deviation $(Ses.Std_{RT})$, and maximum $(Ses.Max_{RT})$ and minimum $(Ses.Min_{RT})$ of the reaction time $RT_{t,i}$ are calculated over all utterances in each session $(t \in T)$. The mean and standard deviation of $Ses.Mean_{RT}, Ses.Std_{RT}, Ses.Max_{RT}, Ses.Min_{RT}$ are calculated as well (the mean is $M.Ses.*$, and the standard deviation is $S.Ses.*$). The dimension of the space is 16.

Speaking length: The statistical parameters of the speaking length $SL_{t,i}$ are calculated for all utterance in all sessions and for each session separately in the same manner using the RT. The dimension is also 16.

Speaking time (Num. of utterance): The total number of utterance ST in all sessions is calculated . The mean and standard deviation of the $ST(i)$ are calculated per session.

B. Indoor daily activity

The indoor location data obtained using the mobile beacons captured how often the participants stayed in each room daily. Wang et al. [23] reported that both social interaction and intellectual stimulation may be relevant to preserving mental function in the elderly through a longitudinal population-based study. From this finding, we hypothesize that the condition of cognitive impairment is associated with the participants'daily activities. It is assumed that participants who stay in the shared space or in others'rooms often are more socially active. Therefore, we calculated the time during which each participant stayed in each room using the location data.

1) When an AP receives a signal from a beacon (ID of the participant), a data sample is added into the database. The data is composed of attributes: (1) ID: M of the AP, (2) ID: P of the beacon, and (3) RSSI: the $RSSI$ of the signal. Each AP: M has a coordinate of position $Pos_M = (X, Y)$ in the residential facility.
2) Every participant lives in one of the two residential facilities. Therefore, we need to extract common activity features which are independent of the residential facility. Common features in both residential facilities are the shared space and the private rooms for participants. We classify the positions, Pos, into three types: participant's own room $(C = 1)$, the shared space (shared living and recreation room) $(C = 2)$, and other places including others' rooms $(C = 3)$ in the residential facility (an example is shown in Figure 1).
3) The place where participant P stays for H hours is estimated using the RSSIs. Let the RSSI vectors of all the Bluetooth signals detected by AP in place $C \in \{1, 2, 3\}$ be $RSSI_{H,C} = \{rssi_{C,1}, \ldots, rssi_{C,N_{H,C}}\}$, where $N_{H,C}$ corresponds to the number of times the signals were received by the APs in place C within H hours. $rssi$ [dBm] is a negative value $(RSSI > -100)$, and we normalize the RSSI by $w_{C,n} = 100 + rssi_{C,n}$.
4) The probability that participant N is staying in place C for H hours is $Pr_{H,C} = \sum_n^{N_{H,C}} w_{C,n}/Z$, where $Z = \sum_C \sum_n w_{C,n}$.

We calculate $Pr_{24,C}$ as the probability of the participant staying in a place for one day (24 h). $Pr_{24,C}$ is calculated for each day for all participants. (ii) in Figure 1 shows an example of the place estimation procedure from the RSSI vectors of the APs. The number of recorded days per participant is distributed from 5-69.

The time s/he stays in place 1-3: The mean and standard deviation of $Pr_{24,1}$ (ratio of time in own room), $Pr_{24,2}$ (ratio of time in shared space), and $Pr_{24,3}$ (ratio of time in other places) are calculated. In addition, we calculate the mean and standard deviation $Pr_{24,2} + Pr_{24,3}$ that correspond to the ratio of time spent in places other than the participant's room. The dimension of the activity feature space is 8.

C. Feature analysis correlated to HDSR

We conducted independent-sample t-tests between multiactivity features and the high or low groups of the dementia scale. In the analysis, the multiactivity features are extracted from all data samples observed during whole term of data collection. Table II shows the results of the t-test for the multi-activity features. We list the features which are significantly different $(p < 0.1)$ between the high/low-score groups. The third row shows the p-values, and fourth row denotes the signs ($>$ or $<$) of magnitude between the high/low groups.

[4] https://webrtc.org/native-code/development/prerequisite-sw/

TABLE II
FEATURES THAT ARE SIGNIFICANTLY DIFFERENT ($p < 0.1$) BETWEEN THE HIGH/LOW SCORE GROUPS OF THE DEMENTIA SCALE

Interaction		p-value	L H
Reaction time	$Max.$	0.059	$<$
Speaking length	$Max.$	0.034	$<$
	$p_{75} - p_{25}$	0.059	$<$
	$M.Ses.Min$	0.034	$<$
	$S.Ses.Mean$	0.049	$<$
	$S.Ses.Min$	0.015	$<$
Speaking time	$Mean$	0.008	$<$
	$M.Ses.Mean$	0.054	$>$

Location		p-value	L H
	std. of $(Pr_{24,3})$	0.090	$>$
	mean of $(Pr_{24,2} + Pr_{24,3})$	0.070	$<$

Concerning the interaction features, the maximum reaction time ($Max.$:) is significantly different ($p = 0.059$), and the reaction time of the low-score group is shorter than that of the high-score group. This result, however, is inconsistent with the previous finding in [3]. According to the present results, the reaction time (the feature is named as "gap" in [3]) of dementia group was longer than that of health control group. Exploring the reason why the result is inconsistent is a remaining work. More features of the speaking length are significantly different. These features are the maximum speaking length ([$Max.$]: $p = 0.034$), the percentile difference ([$p_{75} - p_{25}$]: $p = 0.059$), the standard mean of the minimum speaking length per session ([$M.Ses.Min$]: $p = 0.034$), and the standard deviation of the minimum and mean speaking length per session ([$S.Ses.Min$]: $p = 0.015$, [$S.Ses.Mean$] :$p = 0.049$). The speaking length of the high-score group tended to be longer than that of the low-score group. Concerning activity features, the standard deviation of the ratios of time spent in other places ([std. of $(Pr_{24,3})$]: $p = 0.090$) and the mean ratio of time spent in places except the participant's own room ([mean of $(Pr_{24,2} + Pr_{24,3})$]: $p = 0.070$) are significantly different. This means that the participants in the low-score group tended to stay in their own room for a longer time. Furthermore, this result aligned with the finding that healthy elderly people tend to be social and active.

V. EXPERIMENT

We evaluated the binary classification accuracy of the dementia scale. The goal of this experiment is to answer the following questions: (1) can a model trained using turn-taking features or activity features classify the dementia scale with higher accuracy? (Section V-B) and (2) is fusing turn-taking and indoor activity features effective in improving the classification accuracy? (SectionV-C)

A. Experimental setting

In the experiment, we developed training and test datasets by sampling: (1) S_{sub} sessions from total $\mathbf{S_i}$ sessions that the participant talked with robot and (2) location data of D_{sub} days from the total $\mathbf{D_i}$ days that the sensor data was obtained. The sampling procedure of S_{sub} sessions is as follows: the m th sample of a participant is composed as feature set: x which is extracted using data from the $1 + (m - 1)st$ session to the $S_{sub} + (m - 1)st$ session. st is the step width parameter, which controls the sampling interval, and we set $S_{sub} = D_{sub} = 7,\ st = 3$ in this experiment. The data labeled x is defined as the binary label data (high/low) of the HDSR of the participant. The interaction and activity data were recorded from 19 participants, respectively. When $S_{sub} = D_{sub} = 7$, the interaction and activity dataset are composed of a total 239 and 130 samples, respectively. Leave-one-person cross validation testing [5] is conducted to evaluate the performance. We use the logistic regression classifier (LoReg), linear support vector machine (L-SVM), and random forest classifier (RF) as the classification models. We also use a gradient tree boosting (XGBoost) optimized based on the XGBoost algorithm [24]. Because the sample size is small, we do not use a nonlinear classifier such as a deep neural network, which requires many training samples.

We normalize the data so that each feature has a zero mean and one standard deviation. The parameters of the SVM are optimized using a nested cross-validation scheme, with C parameter values selected from [0.1,1,10]. The parameters of the random forest are optimized similarly using a nested cross-validation scheme, with the numbers of trees per forest selected from [100, 200, 300]. The number of random samples per tree is set as the square root of the training sample set. The parameters of XGBoost are set to 2 for maximum depth of the tree and to 0.3 for the learning rate eta, and the model is optimized with L2 regularization. The balanced accuracy (mean accuracy of both classes) is used as the evaluation criteria for the classification because the numbers of samples are unbalanced. The majority baseline when all samples are classified into one majority class is set as 50%.

B. Classification with interaction behavior and daily activity

To discuss the effectiveness of the feature group, we trained our models with three sets of features: (1) reaction time features, (2) speaking length features, and (3) all interaction features. Cases (1) - (3) included speaking time features. Table III shows the classification results of these models. Columns 2-4 denote the classification accuracy (balanced accuracy) based on (1) the reaction time and (2) the speaking length. The accuracies of all models based on the reaction time are better than those for models based on the speaking length. The best accuracy is 0.705, given using the random forest model.

Column 4 denotes the accuracies of all features. The best accuracy is 0.598, obtained with the random forest model, which was worse than the accuracy of the model based on the reaction time alone by 0.11 points. This result indicates that the speaking length feature did not contribute to improving the accuracy of this task. Although turn-taking features can be extracted more easily than can audio and visual features, the results show that the reaction time effectively produced classification accuracy concerning the score.

Column 5 in Table III shows the results obtained using the activity features. The best accuracy is 0.664, which was obtained with the linear SVM classifier. Although the accuracy

[5] Samples from each person are used for testing in each round.

TABLE III
BALANCED CLASSIFICATION ACCURACY OF MODELS WITH INTERACTION FEATURES AND ACTIVITY FEATURES

	Interaction features			Activity
	Reaction Time	Speaking Length	All	Features
LoReg	0.503	0.500	0.499	0.632
L-SVM	0.660	0.396	0.572	**0.664**
RF	**0.705**	0.633	0.598	0.654
XGBoost	0.489	0.500	0.484	0.587
Mean	0.589	0.507	0.538	**0.634**

TABLE IV
BALANCED CLASSIFICATION ACCURACY OF FUSING INTERACTION AND ACTIVITY FEATURES

	(1) Interaction	(2) Activity	(3) Fusing
LoReg	0.500	0.750	**0.875**
L-SVM	0.500	**0.750**	0.625
RF	0.500	0.750	**0.875**
XGBoost	0.500	0.625	**0.750**
Mean	0.500	0.719	**0.781**

is lower than that obtained using the interaction features (the highest accuracy is nearly 0.7), the classification performance is stable because similar accuracies are obtained by all machine learning models, and the mean accuracy of 0.634 is higher than that (0.507 - 0.589) of the models with interaction features. The results show that the activity feature is also effective for use in the classification of the score.

C. Fusing of interaction and location features

In this section, we analyze the effectiveness of fusing interaction features with activity features. Although the participants who cooperated in providing interaction data are different from those recording activity data, six participants agreed to record both kinds of data. The samples from these six participants are used as the test dataset for the fusion modeling. The training and test data are sampled using the same sampling method described in Section V-A with the same parameters ($S_{sub} = D_{sub} = 7, st = 3$). The total samples of interaction and activity data in the sets are 239, 130, and the test samples total 52 and 45, respectively.

The unit of measurement (session and day) is different between the interaction and activity data, so the classification is conducted per participant. First, training and testing of the model is performed per sample in the same manner described in Section V. The output (score) of the models is summed within the test samples from one participant for use in deciding the label of the participant. In the fusing phase, the total output (score) is summed for both the interaction and activity model (called the latest fusing method), and it is decided as the label with maximum output score. We evaluate the balanced accuracy when data from the six participants is used for testing in cross validation. Three models, (1) a model containing all interaction features, (2) a model containing all activity features, and (3) a late fusion model of (1) and (2), are prepared to evaluate it. Table IV shows the classification accuracy of the fusing of interaction features with activity features. The best accuracy of 0.875 is obtained from the fusion model using logistic regression and random forest. The accuracies of three models except linear SVM are improved by the fusion strategy, and the highest mean accuracy of 0.781 is found for the fusion model. From this result, the fusing of interaction and activity features has the potential to improve the model accuracy.

VI. Discussion

The limitations of this research and future works are summarized in this section. This study used the score of a cognitive test (HDS-R) as labeled data for machine learning. Because we did not use the diagnosis of dementia by professional staff, it is not clarified whether the model and features are effective in detecting dementia. The main focus of future work is to analyze the proposed feature set relative to the diagnosis of dementia. For the interaction data, turn-taking in quiz interactions are used as the feature set. The participants'behavior in small talk with the robot including the greeting is also an important indicator in detecting dementia. The analysis of the relationship between turn taking and the contents of the talk with the robot, along with more effective turn-taking features will be included in future works. For indoor activity data, we used a simple method to detect the location of a participant. It is hard to estimate the accurate position of a participant with this simple method. Indoor positioning algorithms using the RSSIs are surveyed in [25]. In general, we need to develop large-scale reference point set to estimate accurate positions. More precise location sensing will be included in other future work. It is well known that memory problems affect daily activity. For example, people with dementia often repeat the same activity (going to the kitchen or toilet) over and over. Therefore, developing a dementia-dependent activity detection system such as repetitive activities is an important focus of future work.

VII. Conclusion

This paper investigated the integration of two different approaches to predict the score of a cognitive test for the screening of dementia (Hasegawa revised scale) by capturing (1) participants' interaction behaviors with a humanoid robot and (2) living activities in the residence using ubiquitous sensors. In the experiments, we clarified the effectiveness of the interaction feature set and indoor activity feature set in the binary classification of the dementia scale. The best classification accuracy was obtained with interaction feature set, with a balanced accuracy of 0.70 and that obtained with the indoor activity feature set was 0.66. The experimental results also showed that fusing of the interaction and indoor activity features contributes to improving the accuracy. Future work will focus on analyzing the resulting diagnosis of dementia and the proposed features.

Acknowledgment

The authors would like to thank Saeko Doi and Masakazu Kawauti for supporting participants and managing the HDS-R testing. The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions.

References

[1] *Fact sheets: Dementia*, World Health Organization, December 2017, http://www.who.int/en/news-room/fact-sheets/detail/dementia.

[2] H. Tanaka, H. Adachi, N. Ukita, T. Kudo, and S. Nakamura, “Automatic detection of very early stage of dementia through multimodal interaction with computer avatars,” in *Proc. ACM International Conference on Multimodal Interaction (ICMI)*, 2016, pp. 261–265.

[3] H. Tanaka, H. Adachi, N. Ukita, M. Ikeda, H. Kazui, T. Kudo, and S. Nakamura, “Detecting dementia through interactive computer avatars,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, pp. 1–11, 2017.

[4] T. L. Hayes, F. Abendroth, A. Adami, M. Pavel, T. A. Zitzelberger, and J. A. Kaye, “Unobtrusive assessment of activity patterns associated with mild cognitive impairment,” *Alzheimer’s & Dementia: Journal of the Alzheimer’s Association*, vol. 4, no. 6, pp. 395–405, 2008.

[5] Y. Imai and K. Hasegawa, “The revised hasegawa’s dementia scale (hds-r)-evaluation of its usefulness as a screening test for dementia,” *Hong Kong Journal of Psychiatry*, vol. 4, no. 2, p. 20, 1994.

[6] S. O. Orimaye, J. S.-M. Wong, and K. J. Golden, “Learning predictive linguistic features for alzheimer’ s disease and related dementias using verbal utterances,” in *Proc. Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, pp. 78–87.

[7] V. Boschi, E. Catricalà, M. Consonni, C. Chesi, A. Moro, and S. F. Cappa, “Connected speech in neurodegenerative language disorders: a review,” *Frontiers in Psychology*, vol. 8, p. 269, 2017.

[8] V. Masrani, G. Murray, T. Field, and G. Carenini, “Detecting dementia through retrospective analysis of routine blog posts by bloggers with dementia,” in *Proc. Workshop on Biomedical Natural Language Processing (BioNLP)*, 2017, pp. 232–237.

[9] A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert *et al.*, “Automatic speech analysis for the assessment of patients with predementia and alzheimer’s disease,” *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 112–124, 2015.

[10] A. König, A. Satt, A. Sorin, R. Hoory, A. Derreumaux, R. David, and P. H. Robert, “Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people,” *Current Alzheimer Research*, vol. 15, no. 2, pp. 120–129, 2018.

[11] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, “Linguistic features identify alzheimer’s disease in narrative speech,” *Journal of Alzheimer’s Disease*, vol. 49, no. 2, pp. 407–422, 2016.

[12] E. Aramaki, S. Shikata, M. Miyabe, and A. Kinoshita, “Vocabulary size in speech may be an early indicator of cognitive impairment,” *PloS one*, vol. 11, no. 5, p. e0155195, 2016.

[13] V. E. Sturm, M. E. McCarthy, I. Yun, A. Madan, J. W. Yuan, S. R. Holley, E. A. Ascher, A. L. Boxer, B. L. Miller, and R. W. Levenson, “Mutual gaze in alzheimer’s disease, frontotemporal and semantic dementia couples,” *Social Cognitive and Affective Neuroscience*, vol. 6, no. 3, pp. 359–367, 2010.

[14] C. Magai, C. Cohen, D. Gomberg, C. Malatesta, and C. Culver, “Emotional expression during mid to late- stage dementia,” *International Psychogeriatrics*, vol. 8, no. 3, pp. 383–395, 1996.

[15] K. Asplund, A. Norberg, R. Adolfsson, and H. M. Waxman, “Facial expressions in severely demented patients a stimulus response study of four patients with dementia of the alzheimer type,” *International Journal of Geriatric Psychiatry*, vol. 6, no. 8, pp. 599–606, 1991.

[16] G. Stratou and L.-P. Morency, “Multisense—context-aware nonverbal behavior analysis framework: A psychological distress use case,” *IEEE Transactions on Affective Computing*, vol. 8, no. 2, pp. 190–203, 2017.

[17] M. R. Hodges, N. L. Kirsch, M. W. Newman, and M. E. Pollack, “Automatic assessment of cognitive impairment through electronic observation of object usage,” in *International Conference on Pervasive Computing*. Springer, 2010, pp. 192–209.

[18] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, “Automated cognitive health assessment using smart home monitoring of complex tasks,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 6, pp. 1302–1313, 2013.

[19] D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and R. Helaoui, “Smartfaber: Recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment,” *Artificial intelligence in medicine*, vol. 67, pp. 57–74, 2016.

[20] S. Robben, M. Pol, and B. Kröse, “Longitudinal ambient sensor monitoring for functional health assessments: a case study,” in *Proc. International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), Adjunct publication*. ACM, 2014, pp. 1209–1216.

[21] S. Robben, G. Englebienne, and B. Kröse, “Delta features from ambient sensor data are good predictors of change in functional health,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 4, pp. 986–993, 2017.

[22] K. Kim, D. Lee, J. Jhoo, J. Youn, Y. Suh, Y. Jun, E. Seo, and J. Woo, “Diagnostic accuracy of mini-mental status examination and revised hasegawa dementia scale for alzheimer’ s disease,” *Dementia and Geriatric Cognitive Disorders*, vol. 19, no. 5-6, pp. 324–330, 2005.

[23] H.-X. Wang, A. Karp, B. Winblad, and L. Fratiglioni, “Late-life engagement in social and leisure activities is associated with a decreased risk of dementia: A longitudinal study from the kungsholmen project,” *American Journal of Epidemiology*, vol. 155, no. 12, pp. 1081–1087, 2002.

[24] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proc. ACM International Conference on Knowledge Discovery and Data Mining (KDD)*. New York, NY, USA: ACM, 2016, pp. 785–794. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939785

[25] S. He and S. H. G. Chan, “Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons,” *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

Appendix

Influence of the dataset size on the classification accuracy

We investigated how many data (session) samples are required to maintain classification performance by conducting experiments with changing a and b values $[S_{sub} = D_{sub} = a, st = b]$ of Section V-A. Table V shows the dependency of the accuracy on the number of sampling sessions with L-SVM, RF, and LoReg. Rows 1-2, rows 3-4 and row 5 denote the dependency on the samples of the interaction model, activity model and fusion model, respectively. The accuracy of the model with interaction features is higher than 0.62, except $[a = 10, b = 3]$ for the random forest, and a similar accuracy is obtained regardless of the number of samples. From Table V, almost the same accuracy is also obtained for the activity and fusion models regardless of the number of samples.

TABLE V
Dependency of the accuracy on the amount of samples

$[S(D)_{sub}, st]$		[5, 3]	[5, 5]	[7, 3]	[7, 5]	[10, 3]	[10, 5]
Interaction	L-SVM	0.660	0.753	0.660	0.611	0.636	0.728
	RF	0.707	0.683	0.705	0.744	0.616	0.539
Activity	L-SVM	0.669	0.648	0.635	0.637	0.618	0.644
	RF	0.641	0.630	0.653	0.656	0.645	0.660
Fusing	LoReg	**0.875**	0.625	**0.875**	**0.875**	**0.875**	**0.875**