

UC Davis

UC Davis Previously Published Works

Title

Automatic assessment of functional health decline in older adults based on smart home data.

Permalink

<https://escholarship.org/uc/item/04g9b1vs>

Authors

Alberdi Aramendi, Ane

Weakley, Alyssa

Aztiria Goenaga, Asier

et al.

Publication Date

2018-05-01

DOI

10.1016/j.jbi.2018.03.009

Peer reviewed



Published in final edited form as:

J Biomed Inform. 2018 May ; 81: 119–130. doi:10.1016/j.jbi.2018.03.009.

Automatic assessment of functional health decline in older adults based on smart home data

Ane Alberdi Aramendi^{a,1}, Alyssa Weakley^b, Asier Aztiria Goenaga^a, Maureen Schmitter-Edgecombe^b, and Diane J. Cook^c

^aMondragon University, Electronics and Computing Department, Goiru Kalea, 2, Arrasate, 20500, Spain

^bDepartment of Psychology, Washington State University, Pullman, Washington, 99164, USA

^cSchool of Electrical Engineering and Computer Science, Washington State University, Pullman, Washington, 99164, USA

Abstract

In the context of an aging population, tools to help elderly to live independently must be developed. The goal of this paper is to evaluate the possibility of using unobtrusively collected activity-aware smart home behavioral data to automatically detect one of the most common consequences of aging: functional health decline. After gathering the longitudinal smart home data of 29 older adults for an average of > 2 years, we automatically labeled the data with corresponding activity classes and extracted time-series statistics containing 10 behavioral features. Using this data, we created regression models to predict absolute and standardized functional health scores, as well as classification models to detect reliable absolute change and positive and negative fluctuations in everyday functioning. Functional health was assessed every six months by means of the Instrumental Activities of Daily Living-Compensation (IADL-C) scale. Results show that total IADL-C score and subscores can be predicted by means of activity-aware smart home data, as well as a reliable change in these scores. Positive and negative fluctuations in everyday functioning are harder to detect using in-home behavioral data, yet changes in social skills have shown to be predictable. Future work must focus on improving the sensitivity of the presented models and performing an in-depth feature selection to improve overall accuracy.

Graphical Abstract

Correspondence to: Ane Alberdi Aramendi.

¹Telephone: +34 647-504-215

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Keywords

Functional health; Smart home; Activity recognition; Automatic assessment; Behavior; Older adults

1. Introduction

Increasing life expectancy is causing a general aging of the population. As a result, there is a current need to develop systems aimed at early detection of diseases and health issues associated with aging. One consequence of abnormal cognitive aging is the loss of functional skills [1, 2]. Therefore, there is also a current need to create tools and technologies to help the elderly live independently. The current study evaluated the use of unobtrusive sensor technology collected in older adults' homes to automatically assess overall functional health. The term "automatic" implies that data is collected unobtrusively in real time, with no user input (e.g., no buttons to push, no test questions, etc), and treated with specific algorithms to extract useful information from it.

Currently, daily functioning in older adults is primarily assessed through self-report and informant-report questionnaires [3]. Self- and informant-report prove advantageous because these questionnaires are easily administered and considered reasonably accurate given that raters have the opportunity to consider multiple observations of activities performed over periods of time in the real-world. The main disadvantage, however, is bias can be introduced by the reporter for several reasons including lack of insight or awareness, not being present to capture all behavior changes, and the intrinsic tendency to answer questions in a certain manner [4, 5, 6]. Furthermore, raters may fail to recall pertinent information. Alternatively, performance-based assessments that simulate everyday activities in the laboratory are beneficial because they provide objective, quantifiable, and norm-referenced measures of functional ability. However, a major drawback to these assessments is that they take the person out of their natural environment, modifying their usual behavior as a result and missing compensatory strategies that they might be applying in their daily life [7, 8]. Arguably, the ideal strategy to accurately and reliably capture functional decline is to observe daily behavior of individuals where they spend most of their time: at home.

Technology to unobtrusively and ubiquitously monitor peoples' in-home behavior is already available as smart homes [9]. Smart homes represent a useful infrastructure to continuously monitor older adults' behavior in a completely transparent way, gathering real-life data throughout the day and therefore overcoming the main disadvantages of the usual assessment methods. The collected data and machine learning-generated activity labels can

provide a complete view of older adults' behavior in a real-world environment, improving the efficiency and ecological validity of the resulting functional health assessments [10].

Smart home-based behavioral data have already been found to be useful in assisting the elderly in several ways. On one hand, feasibility of systems that use smart home behavioral data to aid in independently living has been demonstrated. For example, prompting technologies designed for elderly with mild cognitive impairment (MCI) [11] or Alzheimer's disease [12] have been developed and tested cross-sectionally in smart home testbeds. On the other hand, longitudinal monitoring of smart home-based behavioral data has shown to be useful to monitor older adults' health state as well as the onset and progress of some age-related diseases and disorders. The overall cognitive ability of older adults has been predicted by unobtrusively collecting in-home behavioral data [13, 14], and more importantly, diseases like MCI [15] and dementia [16] have also been found to correlate with smart home-based behavioral data. Assessment of the psychological health of older adults has also been in the spotlight of some research, confirming the possibility of detecting depression, emotional states [17] or even loneliness [18] of older adults by analyzing their behavioral data. Other overall health predictors such as physical activity have also been assessed by means of such data [17].

Nonetheless, the potential of unobtrusively collected in-home behavioral data to assess older adults' functional health is yet to be analyzed. In this work, we hypothesize that functional difficulties can be detected using unobtrusively collected smart home behavioral data. To verify our hypothesis, we aim to create prediction models for functional health as measured by the Instrumental Activities of Daily Living-Compensation (IADL-C) scale [19] using a longitudinal activity-labeled smart home dataset. We also aim to evaluate performance of the prediction models, as well as selection of behavioral features that contribute the most to IADL-C data prediction. The signal processing approach followed in this work is based on the computation of temporal statistics measuring change in the behavior of the older adults. For that purpose, the Clinical Assessment using Activity Behavior (CAAB) algorithm, which has already been validated in another work for the automatic assessment of cognitive and mobility skills of older adults [20], has been used. Unlike most work in the literature that makes use of group data and absolute behavioral patterns, in this work inter-subject variability is reduced by computing behavioral characteristics separately for each participant. In turn, it also allows to take into account the temporal nature of functional health changes. This approach has not been tested for the detection of daily function decline yet. In fact, we believe that this is the first work aiming at predicting functional health of older adults as measured by the IADL-C scores using unobtrusively collected smart home behavioral data. Furthermore, this work introduces standardization techniques based on a Reliable Change detection to spot and detect time-periods of significant functional change in the older adults. Our study affirms that unobtrusively collected behavioral data can be useful to automatically assess the daily functioning skills of older adults as measured by the IADL-C questionnaire, as well as to detect reliable changes in functional health.

2. Methods

2.1. Data collection

In collaboration with the Center for Studies in Adaptive Systems (CASAS) and the Neuropsychology and Aging Laboratory at Washington State University (WA, USA), we had access to the unobtrusively collected in-home behavioral data of 40 older adults living in 38 smart homes (2 of which were inhabited by two people), as well as to their biannual functional health assessment data. The smart homes used in this study were common apartments enhanced with passive infra-red (PIR) presence sensors. The number of sensors installed in each apartment differed (mean number of sensors installed were 16.52 with a standard deviation of 4.53, ranging from 11 to 26 sensors) depending on the size and shape of the house, but were in all cases strategically placed in specific locations of the houses, including, on top of kitchen devices (stove, sink and refrigerator), office and living room chairs and the bed, as well as installed in the ceiling of different rooms covering the whole room area (e.g. living room, bathroom, dining room, kitchen, laundry, office, bedroom, corridors, etc.). These sensors tracked the movements and activities of the inhabitants by triggering raw sensor-data streams every time a sensor event was detected inside an apartment.

Functional health assessments were collected through the IADL-C questionnaire [19] developed for the early detection of functional deficits and use of compensatory strategies in older adults. The questionnaire assesses IADLs across a number of everyday domains, including phone use, traveling, shopping, cooking, medication management, finances, communication, organization, and social functioning patterns of the participants. As detailed in the IADL-C psychometric paper, a factor analysis grouped the 27-item IADL-C questionnaire into four factors representing different functional abilities: (1) money & self management, (2) home daily living, (3) travel & event memory, and (4) social skills. The four factor analysis derived factors and their respective functional description, Spearman correlation test-retest reliability coefficients, and standard deviations are presented in Table 1. A total “Global Functional Health” score including all four factors is also included.

Smart home sensor data was collected continuously for the duration of the study, which took place from 2011 to 2016, with data collection ranging from < 1 month to 60 months (mean (M) length of the data collection process among the different apartments was 19.95 months, with a standard deviation (SD) of 17.98 months). For the following analyses, data coming from homes with multiple persons were removed (N=2), due to difficulties estimating each individuals activity level. Subjects who had no functional health assessment data (N=2) or who had less than 6 months of behavioral data collected (N=5) were also removed. Therefore, the final dataset contained the behavioral and functional health assessment data of 29 older adults who were living independently and alone in their own smart home residences (M=26 months, SD=17.5 months, range=6–60 months).

2.2. Preprocessing

2.2.1. Day-level behavior feature extraction—The smart home data were a collection of raw sensor-data streams, which collected all sensor events that took place in each

residence during the study period, along with their specific timestamps, sensor IDs and type of event (activation/deactivation). To make the raw sensor-data streams interpretable, we first applied the AR activity recognition algorithm specified in [20], which assigned a specific activity to each sensor entry. This algorithm applies an adaptive length sliding window to the raw sensor data stream to map each one of the sensor events to a value from a pre-defined set of activity labels in real-time. The predefined set of activities consists of specific basic (such as walking or sitting) and instrumental activities of daily living (ADLs) (e.g., cook, eat or personal hygiene activities). This approach takes into account contextual information (such as the activity performed in the previous time-window) in addition to the actual sensor events that fall within the window when identifying the activity being performed.

Accumulated sensor events in a window, as well as time of first and last sensor events, temporal span of the window and mutual information-based influences of all other sensors on the sensor generating the event to be labeled are used as predictors. Three-fold cross validation testing of an activity model learned with this AR algorithm has shown an accuracy exceeding 98% on 30 testbed smart homes in a previous work.

Once the activity-level information was available, we computed 10 daily behavior features for each subject. Python scripts were created for this purpose. The computed day-level activity-features are shown in Table 2.

In order to estimate the daily distance that the subjects were traveling inside their homes, we first created sensor mapping-files based on the floor plan and sensor layout for each residence (see example in Figure 1), where the x-y coordinates of the motion sensor's positions were specified. For 3 of the apartments, we did not have specific information of the positioning of the sensors nor of the distribution within the houses: in these cases, the positions of the sensors were estimated by considering the apartments to be of a similar shape to the rest and checking the activation order of the sensors in the raw sensor data files. Once the positioning of the sensors was specified, we estimated the total walking distance traveled by the inhabitants. For that purpose, we assumed that the inhabitants walked in a straight line from the coverage area of one sensor to the coverage area of another sensor, activating them when they come to be under their same position. Then, we computed the Euclidean distances between randomly-selected locations within the coverage areas of the consecutively activated motion sensors using their x-y coordinates, and we sum all the distances between the sensors activated throughout the day to obtain the daily total walking distance. Note that this approach does not take into account the existence of walls or other obstacles between the sensors, so it just provides an approximation of the real covered distance.

2.2.2. Between-assessments behavior statistics' computation—Once daily activity features for each subject were computed, we used the Clinical Assessment using Activity Behavior (CAAB) [20] algorithm to extract the behavioral statistics of each between-assessment period. RStudio for R [21] was the selected environment for this purpose.

The CAAB algorithm has been introduced in [20]. In brief, each subject's between-assessment daily behavior data were taken and five summarizing time-series statistics were

computed for each behavioral feature of Table 2 in this period: variance, skewness, kurtosis, autocorrelation and change. Because standard assessment was performed every six months, these statistics represent the behavior observed in a smart home for a six month period ending at the assessment date. For this purpose, a log transformation and a Gaussian detrending was first applied to each time-series (behavioral variable) and then the changing time-series statistics for each variable were computed by means of a sliding window of length 7 days. The average of each time-series statistic for the 6-month period was computed and was used for the final predictions. This process can be seen in Figure 2. The resulting preprocessed dataset was a collection of 50 (5 time-series statistics of 10 behavioral features) biannual summary behavior statistics of length $24.0 \pm 13.68(SD)$ months.

2.2.3. Functional health scores' set-up—Our object is to create prediction models that map smart home-based behavior features to health assessment values. In this study our target variables are the IADL-C total and subscore values self-reported by the participant at the end of each corresponding 6-month period.

Self-reported questionnaires can be highly subject-dependent for several reasons. In order to take into account the inter-subject variability that each subjects' age, gender, education or habits might provoke in the scores, we also considered the use of standardized scores for each one of the IADL-C scores for each subject. The standardized scores were computed as the percent change in relation to their baseline values. Baseline IADL-C scores were collected at the first testing session just prior to the beginning of behavioral monitoring with the sensors. The standardized scores were computed as:

$$IADLscore_{std}(i) = \frac{IADLscore(i) - IADLscore_{baseline}}{IADLscore_{baseline}} * 100 \quad (1)$$

Equation 1: Standardized self-reported assessment score at time-point i , computed successively for $i=0$ (baseline), 2, ... I (last assessment point).

With the objective of determining if there was an absolute change in participants' functional health assessment scores both compared to their baseline values ($RCI_{baseline}$) and to the previous assessment point ($RCI_{consecutive}$), we computed the Reliable Change Indexes (RCI) for our IADL-C scores as defined by Christensen and Mendoza [22]. The RCI verifies that the difference between the scores under comparison is greater than a certain level discarding changes that might have appeared due to other reasons such as measurement unreliability. In order to calculate the RCIs for the total IADL-C score and the four IADL-C factors shown in Table 1, we gathered test-retest reliability (r_{score}) and standard deviations (SD_{score}) that the test has shown in its development cohort [19], as shown in Table 1. The RCIs for each subject were thus computed as:

$$RCI_{baseline}(i) = \frac{Score_i - Score_{baseline}}{\sqrt{2SEm}} \quad (2)$$

Equation 2: Reliable Change Index from baseline to assessment time-point i , computed successively for $i=0$ (baseline), 1, ... I (last assessment point).

$$RCI_{consecutive}(i) = \frac{Score_i - Score_{i-1}}{\sqrt{2SEM}} \quad (3)$$

Equation 3: Reliable Change Index between assessment time-points i and $i-1$, computed successively for $i=1, 2, \dots I$ (last assessment point).

where SEM or Standard Error of Measurement represents the expected variation of the observed test scores due to measurement error and is computed as $SEM = SD_{score} \sqrt{1 - r_{score}}$, r_{score} is the test-retest reliability measuring the consistency of the test-scores over time, $Score_i$ is the test score at assessment point i , $Score_{baseline}$ is the test score at the first/baseline assessment and $Score_{i-1}$ is the test score at the previous assessment point.

Therefore, we assigned two new labels to each smart home behavior data instance for the total IADL-C score and each of the four factor subscores. These labels indicate whether the subject suffered a significant change in his/her global functioning and in specific tasks compared with both the baseline assessment and the previous assessment point. This results in a total of 10 labels for each data instance.

Finally, to test the potential of activity-labeled smart home-based behavioral data to detect improvement or decline in everyday functioning, for each subject's IADL-C total score and subscores we computed the difference between each consecutive assessment point. Then, we labeled as positive all the data instances where the subjects self-reported improved everyday functioning (> 0) on the IADL-C while we labeled as negative the behavioral data instances where the subjects self-reported a decline (< 0) in everyday functioning. Thus, five new labels for each behavior data instance are derived from this last step.

We will use machine learning algorithms to learn mappings from the feature vectors to each of these 15 target classes, as well as to predict self-reported IADL-C scores and their standardized versions.

2.3. Functional Health change prediction

The preprocessed dataset resulting from the previous steps was analyzed using Weka [23]. For the four different types of scores which have been introduced in 2.2.3, regression and classification algorithms were built and evaluated, depending on the nature of the scores' data (numeric or nominal labels).

2.3.1. Regression Analyses—First, we performed a regression analysis between the functional health assessment scores and smart home based behavioral data, both for the absolute IADL-C scores and the standardized values. For this purpose, several regression algorithms were implemented using all the behavioral statistics achieved in the previous step and were validated for the prediction of each one of the available IADL-C scores. A 10-fold

cross validation (CV) was used for validation purposes, as well as a leave-one-subject-out cross-validation (LOSOCV) for the absolute scores' case. In the case of LOSOCV, we repeatedly train the model using data for $n-1$ subjects and test on data for the held-out participant (subject n), repeating the process n times and reporting the average of the performance results. We compared the results obtained with the following algorithms: Linear Regression, Linear Support Vector Regression (SVr), SVr with a Radial Basis Function (RBF) kernel, M5 Rules Regression and k Nearest Neighbours (kNN).

2.3.2. Classification Analyses—We then created detection models for the Reliable IADL-C changes using several classification algorithms: AdaBoost, kNN, Linear SVM and Multilayer Perceptron (MLP). The algorithms were trained and validated following a 10-fold cross validation, as well as with a LOSOCV. This process was repeated both for the whole set of behavioral features gathered in the smart homes, and for task-specific behavioral features: sleep-related features, overnight features, mobility, mobility and outing patterns and cooking and eating habits. Table 3 shows the features considered for each task-specific analysis. As a reliable change in prediction scores might be considered to be a rare or unusual event, common classification algorithms might be biased towards the majority class. However, detection of the reliable change event may be the main goal for many applications. To boost detection of these rare events, we tried two approaches that might be more suitable for such unbalanced classification problems: (1) a one-class linear SVM algorithm and (2) the previous algorithms trained with SMOTE-based [24] oversampled datasets. While the former relies on only using minority-class data instances for model training, the latter consists of adding synthetically-created minority-class instances, yielding more class-balanced datasets for training purposes. A rejection rate of 0.1 was used for the one-class linear SVM, which was the empirically selected value in a preliminary test on Reliable baseline total IADL-C change detection. SMOTE algorithm was used to oversample the number of reliable change instances of the original datasets in order to ensure a proportion of at least 40–60% between the two classes. Finally, we aimed at creating prediction models for the daily functioning improvement and decline between consecutive assessment points. For this purpose, we added a fifth classifier to the previous ones, the C4.5 decision tree algorithm. We trained and validated the five classification algorithms using the labels indicating a positive or negative change in these skills.

2.3.3. Evaluation—For all the aforementioned regression and classification models, corresponding pairwise random algorithms were built and evaluated following the same process. The random algorithms provided a basis of comparison to ensure that performance results are not due to chance. These random algorithms were built using a uniformly distributed random data-matrix of the same size as the real behavioral data, while respecting each variable's data range as in the original dataset. The smart home algorithms' performance was compared to their homologous random classifiers' performance to search for statistically significant improvements using smart home based behavioral data. For this purpose, a corrected paired t-test was used. In case of SMOTE-based classifiers, a single run of the algorithms was available, and therefore, a McNemar's test was performed to search for statistically significant improvements compared to the corresponding pairwise random classifiers.

The selected metrics for the regression analyses were the correlation coefficients (r), Root Mean Squared Errors (RMSE) and Mean Absolute Errors (MAE) that compare the actual scores' values and the predicted values using the alternative models. In case of the classification problems, we compared the accuracy (Acc.) and weighted F-scores of the cross-validated results. This last metric was selected to overcome the biased impression that the accuracy can give about a classifier in face of an imbalanced dataset. Therefore, we consider that a certain set of features has prediction ability for the posed classification problem if a t-test shows enough statistical significance indicating that the actual classifier's accuracy or F-score beats the corresponding pairwise random classifier. In case of reliable IADL-C change detection, the cost of missing a true positive might be considered to be higher than having a false positive depending on the application. Equally, the detection of a decline in functional health between assessments might be more important than the detection of an improvement in functional health. Therefore, we also analyze the sensitivity (Sens.) of the smart home-based algorithms to evaluate their ability to predict these events of interest.

Figure 3 gives an overview of the whole research procedure followed in this paper.

3. Results

3.1. Regression Analyses

Table 4 shows the results of the regression algorithms developed using all the behavioral features for the absolute IADL-C test scores, while Table 5 shows the regression results for the standardized IADL-C scores. There is more statistical evidence for the absolute test scores to be predictable with activity-labeled smart home data, and overall, correlations between the actual test-scores and the predicted values from the algorithms are higher in this case than in the case of standardized test scores. When comparing regressors, the SVr algorithm with a RBF kernel worked the best for prediction of the absolute IADL-C scores, achieving a statistically significant prediction in all five cases. Other algorithms have also asserted the possibility of making these predictions, mostly for the total IADL-C scores and the F3 and F4 subscores. In the case of the standardized scores, smart home data contributed to the prediction of all five IADL-C scores, but the largest effect is seen in the global functional health score.

Table 6 shows the LOSOCV regression results for predicting the absolute IADL-C score and subscores using all behavioral features. As expected, correlations between the actual and predicted IADL-C scores are greatly reduced, suggesting the increased difficulty of creating valid general models and the importance of including personal information to adapt the models to each subject.

3.2. Classification Analyses

3.2.1. Reliable change detection—Table 7 shows the results of the classification algorithms for the reliable IADL-C change detection using all the behavioral features gathered in the smart homes. In this case, the kNN and linear SVM algorithms showed to be the most useful, as the former demonstrated statistically significant improvement compared to random classifiers for reliable change detection of F4 from the baseline, while the latter

demonstrated detection power for changes in the total score and F3 subscore from the baseline. Overall, we can appreciate a lack of sensitivity for the positive reliable change detection, but the AdaBoost classifier did perform superior to a random classifier for the detection of a consecutive reliable change in the total IADL-C scores. Results suggest that change in IADL-C scores from baseline are easier to detect than changes between consecutive assessment points.

Table 8 shows the results of LOSOCV classification of RCI change detection using all behavioral features. Overall, results are slightly decayed, but there is still some statistical evidence of improved performance compared to random classifiers. These results suggest that the computation of the reliable change in IADL-C scores is a good way to standardize the values and that this approach can be used to create models for the general population.

Table 9 shows the results classifying reliable change detection using task-specific features. These results suggest that not all of the tasks contribute in the same way for reliable IADL-C change detection: Specifically, cooking and eating patterns are useful in this study for the detection of the total and F4 subscores. The total score, F3 and F4 subscores have shown to be detectable by mobility and outing patterns while sleeping and overnight patterns are related to the changes in the total IADL-C scores and F1 and F4 subscores. Interestingly, mobility features and the combination of mobility and outing patterns showed to be useful for the applications where we are more interested in reliably detecting the change in global IADL-C scores, as their contribution to the sensitivity of the classifiers has shown to be statistically significant for three of the experiments. Sleep-related features have shown to be contributive for changes in F1 subscore from baseline. While kNN and MLP were found to be the best algorithms for the prediction models in this case, we can specially notice the biased behavior and lack of sensitivity of the linear SVM models. Finally, the results of employing activity-specific features in LOSOCV evaluation, shown in Table 10, verify the validity of RCI scores to create inter-individual models based on smart home data.

3.2.2. Sensitivity improvement—Table 11 shows the results of reliable IADL-C detection using the one-class linear SVM algorithm. Overall, an improvement in terms of sensitivity can be appreciated compared to the results obtained with other classification algorithms at the expense of accuracy and F-score values. These algorithms show a higher number of false alarms, and therefore they might be only useful when detection of the reliable change is critical.

Finally, Table 12 shows the results for reliable change detection using all behavioral features and SMOTE-based oversampled datasets for training purposes. As shown, sensitivity of the models is improved compared to the initial models, at the expense of precision. Nonetheless, some of these results maintain a favorable trade-off between sensitivity and overall performance of the classifiers, overcoming the barrier of 60% accuracy and even 70% for sensitivity, and thus, can be very interesting for automated functional health assessment. The kNN algorithm yields improved performance in comparison with random-data based algorithms for all IADL-C scores, while AdaBoost, linear SVM and MLP algorithms also yield statistically improved performance for total, F2, F3 and F4 subscores.

3.2.3. Positive/Negative change detection—Table 13 shows the results of the classification algorithms for the detection of positive and negative changes in IADL-C total score and subscores between consecutive assessment points. In this case, the C4.5 algorithm has shown enough statistical significance to accept that we are able to detect the improvement and decline of participants' social skills (F4 subscores) from smart home data, while kNN algorithm has shown increased sensitivity for the detection of decline in overall daily functioning.

4. Discussion

The problem addressed in this paper is highly challenging: Specifically, our aim was to predict older adults' functional health from unobtrusively collected behavioral data inside their own apartments. Despite the difficulty of the task, our results have demonstrated the possibility of predicting functional health and changes in everyday functioning from activity-labeled smart home data.

Although we could have assumed that subscore F2, reflecting home daily living, would be the most correlated score to the behavior data, regression analyses proved this to be false. The absolute total IADL-C score, as well as the F1 and F3 subscores, which reflect the money/self-management skills and travel/event memory abilities, appeared most related to the sensor behavior data. In prior work [19], informants reported that individuals with mild cognitive impairment experienced the greatest changes in the money/self-management domain followed by the travel/event memory domain. Therefore, early identification of functional difficulties in these domains based on sensor data could be of importance for early intervention. In addition, all of the IADL-C scores were predicted from unobtrusively collected behavior data with statistically significant performance. Furthermore, absolute scores were more predictable than the standardized ones, suggesting that IADL-C scores are directly comparable between subjects. The importance of adding personalized data points to the models was also demonstrated, as correlations obtained with a LOSOCV were shown to be much lower than using 10-fold CV, where data points from the same people might be used to both train and test the models. This suggests that general models would highly benefit from a system that could actively learn individuals' behavior and functional health state, increasing their prediction performance while collecting data by being adapted to each user.

In this paper, reliable change in everyday functioning was predicted both compared to baseline and between consecutive assessment points. Nonetheless, we cannot forget that we are using the same behavioral data points for both cases, and that these data points are coming from each between-assessment period. This means that time-series statistics of the behavioral features collected in each between-assessment period not only help in predicting change in the corresponding period but also change compared with baseline. Reliable changes in the total IADL-C scores, both compared to the baseline and compared to previous assessment points, have shown to be predictable from smart home data.

Unexpectedly, we also noticed a high predictability of F4 IADL-C subscore, which is related to social skills of older adults, from in-home data. Even if social skills of the elderly might

be something that can be much better appreciated when they are outside, where they relate to other people and they carry-out social activities, we have demonstrated that there are some in-home characteristics that might also help in predicting these abilities. The combination of in-home mobility and outing patterns have shown to be related to social skills, as well as cooking and eating habits. This is indeed logical, as the more an elder is leaving his/her home, the more social life they are likely engaging in. This agrees with previous work [17] which has also reported an association between increased time spent out of home and better social health (in terms of decreased loneliness and better mood). Not only that, but sensors might also be detecting more variability in eating patterns of the elderly who have the best social skills because they might be going out for meals more frequently. This could also explain the correlation found between the overnight patterns and social skills, as going out more and having a less routine life might increase the variability in their nighttime behavior. A LOSOCV has shown results of similar magnitudes as the 10-fold CV, suggesting the possibility of creating inter-individual models based on this approach.

For the functional health improvement and decline detection, results have been more moderate. This is understandable, as this problem has an added difficulty for two reasons. On one hand, we were considering that a change in IADL-C scores occurred at every assessment point compared the previous one, even if this change was not really significant or it was simply inestimable. On the other hand, we were aiming at distinguishing positive and negative changes in IADL-C scores when the time-series statistics that we are extracting from our behavioral data might not necessarily reflect positive or negative change in behavior. Even so, we were able to demonstrate that positive and negative changes in social skills are predictable using smart home data.

Finally, we observe that the use of specific algorithms for imbalanced datasets can significantly help in gaining sensitivity for the reliable change events' detection. In this paper, as in most research focusing on detection of health issues, we face a class imbalance problem. A reliable change in functional health is a rare event, but is likely the event of interest for most applications. Ideally, we would like to have an algorithm with a high sensitivity and high precision or low number of false alarms, but we usually have to seek a trade-off between these parameters. We believe that the algorithms that have been built using SMOTE-based oversampled datasets have shown interesting results in this sense, being more useful than one-class classification algorithms. Reliable changes in total score and F1 and F3 subscores from baseline can be detected with a sensitivity of up to 75% and overall accuracies of 60%, which is not bad for early models and encourages us to keep working further in feature selection and in sensitivity boosting. Significance tests have confirmed that smart home data can be used to predict all five IADL-C scores.

In terms of classification algorithms, linear SVM has shown the least interesting results. In most of the cases, models created using this algorithm are highly biased towards the majority class, showing a null sensitivity for reliable change detection and not demonstrating any statistical significance using smart home data against the use of random data. At the other end is the kNN algorithm, which has pleasantly surprised us in almost all established problems, finding in many cases the best results and the biggest amount of significant

improvements compared to random classifiers. It is certainly an algorithm to consider for future research in problems with similar characteristics.

The work being presented herein is aligned with the current emerging paradigm of the Internet of Things (IoT), which aims at building up a globally interconnected continuum of a variety of objects in the physical environment [25, 26]. IoT has become one of the research priorities in multiple disciplines, including healthcare. The main goal of IoT-enabled healthcare is to design and develop ubiquitous Information and Communication-based solutions for delivering high-quality patient-centered health services. This way, it is intended to propose economically viable alternatives to the traditional healthcare systems in order to mitigate the consequences of the continued aging of the population [27]. Our approach contributes towards this goal by offering an inexpensive ubiquitous monitoring system for the detection of functional-health decline. Taking into account that most people who are part of the largest collective in the developed countries suffer from functional health decline at some point, this work is of great interest for a huge number of potential end-users.

Besides, the system being proposed in this work could be extensible to a wide variety of applications with a little adaptation work, thus expanding its field of use and the list of benefited users. For instance, such an ubiquitous monitoring of people's behavior could be used as a follow-up of a therapy or rehabilitation program in the overall population, improving its efficiency and success, as it could also be used as an overall health monitor. In addition, Emergency Medical Services (EMS) could be improved by automatically detecting in-home emergencies [28]. Moreover, persuasive prompts could be given to the inhabitants based on their behavior in order to guide cognitively impaired people through daily activities [11] or to enhance their emotional [29] and overall wellbeing. Finally, smart hospital services [30, 31] could be deployed by offering a more personalized in-home hospitalization. Nonetheless, there are still some issues that should be addressed in order to implement such a system in real-life. These include lifelogging issues [32], the high volume of generated data or security and privacy issues [33].

5. Conclusions

This work has demonstrated the possibility of detecting functional health decline in older adults from unobtrusively collected in-home behavioral data. We believe that the results shown herein are important, as they suggest the possibility of implementing an IoT-enabled system that can benefit our increasingly older society. The models shown in this paper are early models, which were mainly aimed at demonstrating the feasibility of such a system and providing insight into the behavioral features that might be used for this purpose, more than to create very accurate and likely overfitted models. The results shown in this paper must be completed and improved with more data and algorithmic solutions that might better adapt to the imbalanced detection problems posed herein before their implementation in real-world settings. Therefore, future work will focus on improving the sensitivity of the models without increasing the false alarm rate, by performing a more in-depth feature selection analysis, as well as designing more suitable algorithms for imbalanced datasets and verifying the results in a scaled longitudinal dataset.

Acknowledgments

Research reported in this paper was supported by the National Institutes of Health under award number R01EB015853.

References

1. West, RL. Planning practical memory training for the aged. In: Poon, LW.Rubin, DC., Wilson, BA., editors. *Everyday cognition in adulthood and late life*. Cambridge University Press; Cambridge: p. 573-597. URL <http://ebooks.cambridge.org/ref/id/CBO9780511759390A047>
2. West, RL. Compensatory strategies for age-associated memory impairment. In: Baddeley, AD.Wilson, BA., Watts, FN., editors. *Handbook of Memory Disorders*. John Wiley; London: 1995. p. 481-500.
3. Schmitter-Edgecombe M, Parsey C, Cook DJ. Cognitive correlates of functional performance in older adults: comparison of self-report, direct observation, and performance-based measures. *Journal of the International Neuropsychological Society: JINS*. 2011; 17(5):853–864. arXiv:NIHMS150003. DOI: 10.1017/S1355617711000865 [PubMed: 21729400]
4. Dassel KB, Schmitt FA. The impact of caregiver executive skills on reports of patient functioning. *The Gerontologist*. 2008; 48(6):781–92. URL <http://www.ncbi.nlm.nih.gov/pubmed/19139251>. [PubMed: 19139251]
5. Bertrand RM, Willis SL. Everyday problem solving in Alzheimer’s patients: A comparison of subjective and objective assessments. *Aging & Mental Health*. 1999; 3(4):281–293. URL www.tandfonline.com/doi/abs/10.1080/13607869956055. DOI: 10.1080/13607869956055
6. Richardson ED, Nadler JD, Malloy PF. Neuropsychologic prediction of performance measures of daily living skills in geriatric patients. *Neuropsychology*. 1995; 9(4):565–572. URL doi.apa.org/getdoi.cfm?doi=10.1037/0894-4105.9.4.565. DOI: 10.1037/0894-4105.9.4.565
7. Myers AM, Holliday PJ, Harvey KA, Hutchinson KS. Functional performance measures: are they superior to self-assessments? *Journal of gerontology*. 1993; 48(5):M196–206. URL <http://www.ncbi.nlm.nih.gov/pubmed/8366262>. [PubMed: 8366262]
8. Zimmerman SI, Magaziner J. Methodological issues in measuring the functional status of cognitively impaired nursing home residents: the use of proxies and performance-based measures. *Alzheimer disease and associated disorders*. 1994; 8(Suppl 1):S281–90. URL <http://www.ncbi.nlm.nih.gov/pubmed/8068270>. [PubMed: 8068270]
9. Chan M, Esteve D, Escriba C, Campo E. A review of smart homes - Present state and future challenges. *Computer Methods and Programs in Biomedicine*. 2008; 91(1):55–81. DOI: 10.1016/j.cmpb.2008.02.001 [PubMed: 18367286]
10. Lyons BE, Austin D, Seelye A, Petersen J, Yeagers J, Riley T, Sharma N, Mattek N, Wild K, Dodge H, Kaye JA. Pervasive computing technologies to continuously assess Alzheimer’s disease progression and intervention efficacy. *Frontiers in Aging Neuroscience*. Jun.2015 7:1–14. DOI: 10.3389/fnagi.2015.00102 [PubMed: 25653617]
11. Seelye AM, Schmitter-Edgecombe M, Cook DJ, Crandall A. Naturalistic Assessment of Everyday Activities and Prompting Technologies in Mild Cognitive Impairment. *Journal of the International Neuropsychological Society*. 2013; 19(4):442–452. DOI: 10.1017/S135561771200149X [PubMed: 23351284]
12. Lapointe, J., Bouchard, B., Bouchard, J., Potvin, A., Bouzouane, A. Smart homes for people with Alzheimer’s disease. *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments - PETRA '12*; New York, New York, USA: ACM Press; 2012. p. 1. URL <http://dl.acm.org/citation.cfm?doid=2413097.2413135>
13. Dawadi PN, Cook DJ, Schmitter-Edgecombe M. Automated Cognitive Health Assessment Using Smart Home Monitoring of Complex Tasks. *IEEE Trans on Systems, Man, and Cybernetics – Part C: Applications and Reviews*. 2013; 43(6):1302–1313. DOI: 10.3233/THC-130734
14. Dawadi PN, Cook DJ, Schmitter-Edgecombe M. Automated Cognitive Health Assessment From Smart Home-Based Behavior Data. *IEEE Journal of Biomedical and Health Informatics*. 2016; 20(4):1188–1194. DOI: 10.1109/JBHI.2015.2445754 [PubMed: 26292348]

15. Hayes TL, Abendroth F, Adami A, Pavel M, Zitzelberger Ta, Kaye Ja. Unobtrusive assessment of activity patterns associated with mild cognitive impairment. *Alzheimer's & Dementia*. 2008; 4(6): 395–405. DOI: 10.1016/j.jalz.2008.07.004
16. Galambos C, Skubic M, Wang S, Rantz M. Management of dementia and depression utilizing in-home passive sensor data. *Gerontechnology*. 2013; 11(3):457–468. DOI: 10.4017/gt.2013.11.3.004.00 [PubMed: 24049513]
17. Petersen J, Austin D, Mattek N, Kaye J. Time out-of-home and cognitive, physical, and emotional wellbeing of older adults: A longitudinal mixed effects model. *PLoS ONE*. 2015; 10(10):1–16. DOI: 10.1371/journal.pone.0139643
18. Austin J, Dodge HH, Riley T, Jacobs PG, Thielke S, Kaye J. A Smart-Home System to Unobtrusively and Continuously Assess Loneliness in Older Adults. *IEEE Journal of Translational Engineering in Health and Medicine*. 2016; 4:1–11. DOI: 10.1109/JTEHM.2016.2579638
19. Schmitter-Edgecombe M, Parsey C, Lamb R. Development and psychometric properties of the instrumental activities of daily living: compensation scale. *Archives of clinical neuropsychology: the official journal of the National Academy of Neuropsychologists*. 2014; 29(8):776–92. URL <http://www.ncbi.nlm.nih.gov/pubmed/25344901> DOI: 10.1093/arclin/acu053 [PubMed: 25344901]
20. Dawadi P, Cook DJ, Schmitter-Edgecombe M. Automated clinical assessment from smart home-based behavior data. *IEEE Journal of Biomedical and Health Informatics*. 2015; 99164:1–12. DOI: 10.1109/JBHI.2015.2445754
21. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2016. URL <https://www.r-project.org/>
22. Christensen L, Mendoza JL. A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*. 1986; 17(3):305–308. DOI: 10.1016/S0005-7894(86)80060-0
23. Frank, E., Hall, MA., Witten, IH. Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques”. 4. 2016. The WEKA Workbench.
24. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Int Res*. 2002; 16(1):321–357.
25. Riazul Islam SM, Kwak Daehan, Humaun Kabir M, Hossain M, Kwak Kyung-Sup. The Internet of Things for Health Care: A Comprehensive Survey. *IEEE Access*. 2015; 3:678–708. URL <http://ieeexplore.ieee.org/document/7113786/>. DOI: 10.1109/ACCESS.2015.2437951
26. Acampora G, Cook DJ, Rashidi P, Vasilakos AV. A Survey on Ambient Intelligence in Healthcare. *Proceedings of the IEEE*. 2013; 101(12):2470–2494. URL <http://ieeexplore.ieee.org/document/6579688/>. DOI: 10.1109/JPROC.2013.2262913 [PubMed: 24431472]
27. Qi J, Yang P, Min G, Amft O, Dong F, Xu L. Advanced internet of things for personalised healthcare systems: A survey. *Pervasive and Mobile Computing*. 2017; 41:132–149. URL <http://dx.doi.org/10.1016/j.pmcj.2017.06.018>. DOI: 10.1016/j.pmcj.2017.06.018
28. Lin, Yong, Lu, Xingjia, Fang, Fang, Fan, Jianbo. 2013 First International Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare (Ubi-HealthTech). *IEEE*; 2013. Personal health care monitoring and emergency response mechanisms; p. 1-5. URL <http://ieeexplore.ieee.org/document/6708052/>
29. Suryadevara, NK., Quazi, M., Mukhopadhyay, SC. Smart Sensing System for Human Emotion and Behaviour Recognition. In: Kundu, MK.Mitra, S.Mazumdar, D., Pal, SK., editors. *Perception and Machine Intelligence*, Vol. 7143 of Lecture Notes in Computer Science. Springer Berlin Heidelberg; Berlin, Heidelberg: 2012. p. 11-22. URL <http://www.springerlink.com/index/10.1007/978-3-642-27387-2>
30. Thangaraj, M., Ponmalar, PP., Anuradha, S. Internet Of Things (IOT) enabled smart autonomous hospital management system -A real world health care use case with the technology drivers. 2015 *IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*; *IEEE*; 2015. p. 1-8. URL <http://ieeexplore.ieee.org/document/7435678/>
31. Sanchez D, Tentori M, Favela J. Activity Recognition for the Smart Hospital. *IEEE Intelligent Systems*. 2008; 23(2):50–57. URL <http://ieeexplore.ieee.org/document/4475859/>. DOI: 10.1109/MIS.2008.18

32. Yang P, Stankevicius D, Marozas V, Deng Z, Liu E, Lukosevicius A, Dong F, Xu L, Min G. Lifelogging Data Validation Model for Internet of Things Enabled Personalized Healthcare. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2018; 48(1):50–64. URL <http://ieeexplore.ieee.org/document/7516690/>. DOI: 10.1109/TSMC.2016.2586075
33. Al Ameen M, Liu J, Kwak K. Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications. *Journal of Medical Systems*. 2012; 36(1):93–101. URL <http://link.springer.com/10.1007/s10916-010-9449-4>. DOI: 10.1007/s10916-010-9449-4 [PubMed: 20703745]

Highlights

- A method to automatically assess older adults' functional health was presented.
- IADL-C scale was used to evaluate older adults' functional health.
- Unobtrusively collected in-home behavioral data was used for prediction.
- Change in behavior was analyzed instead of absolute behavior characteristics.
- Inter-subject standardization was done using the Reliable Change Index.

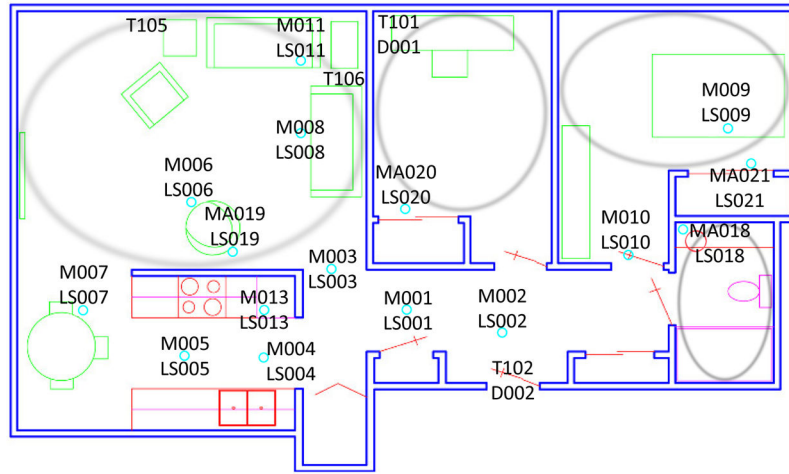


Figure 1.
Floor plan and sensor layout of one of the residences of the study

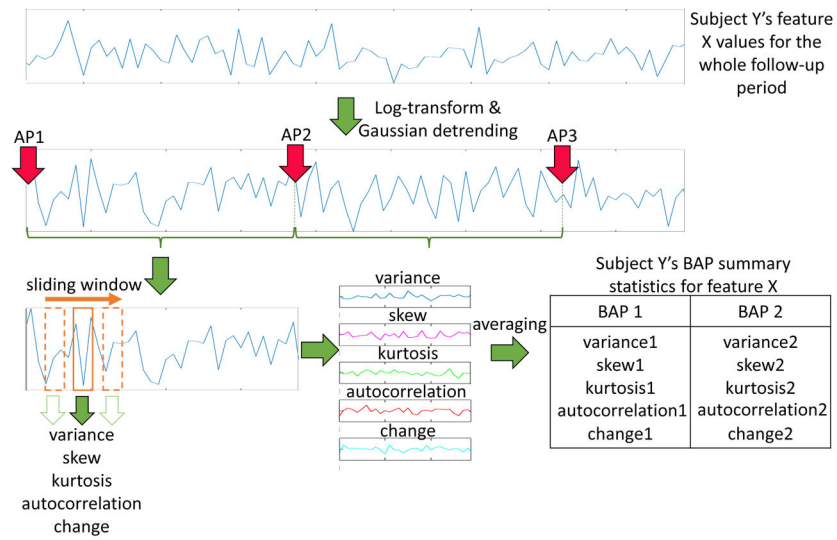


Figure 2. Between-assessment summary statistics' computation (AP: Assessment Point, BAP: Between-Assessment Period)

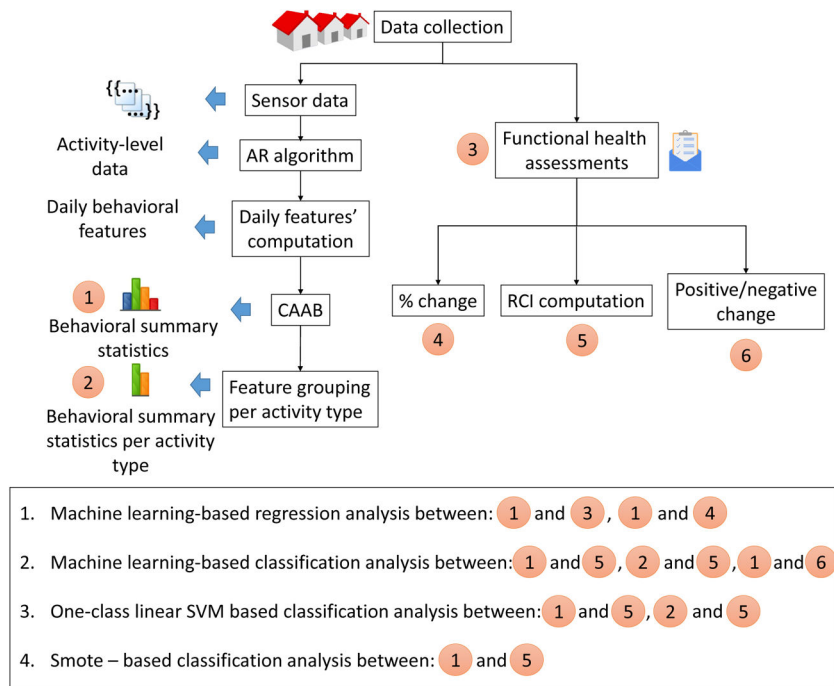


Figure 3.
Flow-chart of the whole method.

Table 1

IADL-C Scores' description, test-retest reliability and standard deviations

Score	Description	<i>r</i>_{score}	<i>SD</i>_{score}
IADL-C Total	Global Functional Health	0.91	1.64
IADL-C Factor 1	Money and self-management	0.91	1.64
IADL-C Factor 2	Home daily living	0.76	1.21
IADL-C Factor 3	Travel and event memory	0.70	1.25
IADL-C Factor 4	Social skills	0.70	1.03

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Day-level activity features included in the study

Type	Day-level features
Duration of specific activities (6 features)	Time spent per day in cooking, eating, relaxing, carrying out personal hygiene activities, being out of home and nighttime toileting activities
Sleep-related (2 features)	The daily sleep duration and frequency
Mobility-related (2 features)	The total number of activated sensors and the total distance covered walking inside the apartment per day

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Task-specific grouping of the daily features

Group	Day-level features
Sleep-related	The daily sleep duration and frequency
Overnight patterns	Sleep-related features + time spent per day in nighttime toileting activities
Mobility-related	The total number of activated sensors and the total distance covered walking inside the apartment per day
Mobility & outings	Mobility-related + time spent per day in being out of home
Cooking & eating	Time spent per day in cooking and eating

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Regression results for the absolute IADL-C test scores using all behavioral features and 10-fold CV.

	Linear Regression			Linear SVr			RBF SVr			M5 Rules			kNN		
	<i>r</i>	RMSE	MAE	<i>r</i>	RMSE	MAE	<i>r</i>	RMSE	MAE	<i>r</i>	RMSE	MAE	<i>r</i>	RMSE	MAE
IADL-C	0.21	21.58 *	15.99 *	0.17 *	22.86 *	15.81 *	0.22 *	20.91	13.58	0.29 *	20.24 *	14.78 *	0.01	32.57	21.01
IADL-C F1	0.14	15.14	11.65	0.13	15.17 *	11.56 *	0.29 *	12.90 *	9.20 *	0.27	12.93 *	10.15 *	0.02	17.61	12.77
IADL-C F2	0.06	6.59	4.93	0.06	6.28 *	4.31 *	0.12 *	5.60 *	3.15 *	0.10	5.76	4.16	0.03	10.32	5.90
IADL-C F3	0.22	4.42 *	3.32 *	0.19 *	4.63 *	3.22	0.26 *	4.26 *	2.55 *	0.23	4.27	3.17	0.02	7.58	4.75
IADL-C F4	0.00	1.69 *	1.03 *	0.18 *	1.62	0.79	0.19 *	1.57	0.65 *	0.00	1.69 *	1.03 *	0.04	1.59 *	0.66 *

* Statistically significant improvement (p<0.05) in comparison to the corresponding pairwise random algorithm.

Table 5

Regression results for the standardized IADL-C test scores using all behavioral features and 10-fold CV.

	Linear Regression			Linear SVr			RBF SVr			M5 Rules			kNN		
	<i>r</i>	RMSE	MAE	<i>r</i>	RMSE	MAE	<i>r</i>	RMSE	MAE	<i>r</i>	RMSE	MAE	<i>r</i>	RMSE	MAE
IADL-C	0.12	36.52	28.50	0.10	33.89*	26.64*	0.11	25.92*	17.97*	0.07	33.96	23.80	0.14	43.14	33.10
IADL-C F1	0.21	46.25	36.57	0.22	45.19	36.40	0.21	31.60	23.99*	0.11	35.78	27.52	0.04	43.67	34.24
IADL-C F2	0.03	42.49	33.00	0.06	34.26	25.96	0.03	28.80	18.56*	0.02	41.37	26.83	0.12	48.52	34.58
IADL-C F3	0.02	80.20	59.63	0.11	65.33	45.15	0.01	53.96	31.76*	0.22*	67.56	44.07	0.07	110.24	66.79
IADL-C F4	0.00	58.32*	35.62*	0.00	58.44	30.19	0.02	54.71	23.35*	0.01	58.21	35.23	0.19*	59.30	25.96

* Statistically significant improvement (p<0.05) in comparison to the corresponding pairwise random algorithm.

Table 6

Regression results for the absolute IADL-C test scores using all behavioral features for LOSOCV.

	Linear Regression			Linear SVr			RBF SVr			M5 Rules			kNN		
	r	RMSE	MAE	r	RMSE	MAE	r	RMSE	MAE	r	RMSE	MAE	r	RMSE	MAE
IADL-C	0.03	17.52	15.20	0.01	17.15*	14.87*	0.02	14.15	12.75	0.02	16.52*	14.58	0.05	27.51	22.41
IADL-C F1	0.02	11.37	10.05	0.01	13.19	11.63	0.12	9.44	8.65	0.06*	11.61	10.20	0.00	15.24	12.95
IADL-C F2	0.08	5.03	4.50	0.13	4.68*	4.07*	0.03	3.34*	2.94*	0.06	4.24	3.81	0.00	8.50	6.51
IADL-C F3	0.07	6.14*	4.82*	0.03	3.16*	2.77*	0.01	2.78*	2.33*	0.02	3.14*	2.82	0.07	6.14	4.82
IADL-C F4	0.03	1.06*	0.91*	0.09	0.93	0.71	0.11	0.82	0.58	0.03	1.06*	0.91*	0.00	0.88*	0.59*

* Statistically significant improvement (p<0.05) in comparison to the corresponding pairwise random algorithm.

Table 7

Reliable IADL-C change detection results with a 10-fold CV using all behavioral features.

Vars	AdaBoost			kNN			LinearSVM			MLP		
	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.
<i>RCI_{baseline} - total</i>	68.46	0.65	0.23	64.61	0.67	0.31	73.59*	0.62*	0.05	59.59	0.63	0.25
<i>RCI_{baseline} - F1</i>	57.83	0.53	0.16	62.11	0.65	0.38	65.77	0.56	0.07	63.27	0.60	0.33
<i>RCI_{baseline} - F2</i>	91.91	0.90	0.00	86.38	0.88	0.00	94.23	0.91	0.00	86.44	0.88	0.05
<i>RCI_{baseline} - F3</i>	81.31	0.77	0.12	67.16	0.68	0.07	84.29	0.77*	0.00	75.44	0.73	0.16
<i>RCI_{baseline} - F4</i>	93.82	0.92	0.00	94.31	0.93*	0.00	95.06	0.93	0.00	92.82	0.91	0.00
<i>RCI_{consecutive} - total</i>	66.94	0.64	0.25*	63.32	0.60	0.20	71.08	0.61	0.07	62.49	0.63	0.26
<i>RCI_{consecutive} - F1</i>	65.27	0.58	0.11	71.33	0.67	0.28	69.87	0.58	0.02	64.92	0.60	0.24
<i>RCI_{consecutive} - F2</i>	87.13	0.84	0.01	80.79	0.81	0.00	90.13	0.85	0.00	81.17	0.83	0.11
<i>RCI_{consecutive} - F3</i>	83.08	0.78	0.00	68.28	0.69	0.03	85.96	0.80	0.00	74.03	0.70	0.01
<i>RCI_{consecutive} - F4</i>	94.23	0.91	0.00	93.56	0.91	0.00	94.23	0.91	0.00	90.17	0.91	0.08

* Statistically significant improvement (p<0.05) in comparison to the corresponding pairwise random algorithm.

Table 8

Reliable IADL-C change detection results with a LOSOCV using all behavioral features.

	AdaBoost			kNN			LinearSYM			MLP		
	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.
$RCI_{baseline} - total$	67.77	0.65	0.22	67.77	0.67	0.31	68.60*	0.61	0.03	57.85*	0.59	0.28
$RCI_{baseline} - F1$	61.16	0.60	0.35	67.77	0.67	0.40	64.46	0.55	0.05	62.81	0.63	0.45
$RCI_{baseline} - F2$	90.08	0.89	0.00	87.6	0.88	0.00	94.21	0.91	0.00	85.12	0.87	0.00
$RCI_{baseline} - F3$	79.24	0.77	0.16	67.77	0.69	0.11	84.30*	0.77	0.00	72.72	0.73	0.16
$RCI_{baseline} - F4$	94.22	0.92	0.00	95.04	0.93	0.00	95.04	0.93	0.00	94.21*	0.92	0.00
$RCI_{consecutive} - total$	66.12	0.62	0.15	63.63	0.61	0.21	67.77*	0.60	0.06	63.63	0.61	0.21
$RCI_{consecutive} - F1$	63.64	0.59	0.11	69.42	0.67	0.29	64.46	0.56	0.00	64.46	0.63	0.29
$RCI_{consecutive} - F2$	88.43	0.85	0.00	83.47	0.82	0.00	90.08	0.85	0.00	80.17	0.80	0.00
$RCI_{consecutive} - F3$	85.12	0.79	0.00	69.42	0.71	0.06	85.95*	0.79	0.00	76.04	0.75	0.06
$RCI_{consecutive} - F4$	94.22	0.91	0.00	94.21	0.91	0.00	94.21	0.91	0.00	89.26*	0.89	0.00

* Statistically significant improvement ($p < 0.05$) in comparison to the corresponding pairwise random algorithm.

Table 9 Results for 10-fold CV classification of Reliable IADL-C change detection using activity-specific features.

Vars	AdaBoost			kNN			LinearSVM			MLP		
	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.
<i>Only cook and eat</i>												
$RCI_{baseline} - total$	70.28	0.62	0.05	49.15	0.50	0.40	73.59	0.62	0.00	65.03	0.61	0.14
$RCI_{baseline} - F1$	60.73	0.54	0.10	63.38	0.59	0.24	66.92	0.54	0.00	62.47	0.59	0.28
$RCI_{baseline} - F2$	92.48	0.91	0.04	89.58	0.89	0.00	94.23	0.91	0.00	91.83	0.90	0.00
$RCI_{baseline} - F3$	84.21	0.77	0.00	55.20	0.60	0.42	84.29	0.77	0.00	78.50	0.74	0.00
$RCI_{baseline} - F4$	95.06	0.93	0.00	94.49*	0.92*	0.00	95.06	0.93	0.00	95.06	0.93	0.00
$RCI_{consecutive} - total$	71.53	0.65	0.18	49.28	0.51	0.45	71.92	0.60	0.00	71.67*	0.68*	0.30
$RCI_{consecutive} - F1$	69.45	0.62	0.09	64.14	0.60	0.18	71.09	0.59	0.00	63.83	0.60	0.18
$RCI_{consecutive} - F2$	88.48	0.85	0.00	85.31	0.84	0.09	90.13	0.85	0.00	87.38	0.84	0.02
$RCI_{consecutive} - F3$	84.88	0.79	0.00	58.28	0.64	0.52	85.96	0.80	0.00	81.41	0.77	0.01
$RCI_{consecutive} - F4$	93.90	0.91	0.00	91.60	0.90	0.00	94.23	0.91	0.00	93.98*	0.91*	0.00
<i>Only mobility</i>												
$RCI_{baseline} - total$	70.64	0.62	0.03	64.74	0.66	0.60*	73.59	0.62	0.00	67.66	0.66	0.34
$RCI_{baseline} - F1$	56.92	0.53	0.16	48.24	0.48	0.41	66.92	0.54	0.00	56.87	0.54	0.25
$RCI_{baseline} - F2$	93.57	0.92	0.11	73.37	0.79	0.01	94.23	0.91	0.00	91.08	0.90	0.10
$RCI_{baseline} - F3$	79.92	0.76	0.07	82.33	0.82	0.38	84.22	0.77	0.00	81.68*	0.80*	0.30
$RCI_{baseline} - F4$	94.81	0.93	0.00	90.92	0.90	0.00	95.06	0.93	0.00	91.35	0.91	0.00
$RCI_{consecutive} - total$	66.81	0.60	0.09	56.08	0.57	0.60*	71.76	0.60	0.00	65.11	0.63	0.31
$RCI_{consecutive} - F1$	65.47	0.58	0.08	47.56	0.49	0.39	70.94	0.59	0.00	62.16	0.60	0.27
$RCI_{consecutive} - F2$	87.32	0.84	0.00	70.58	0.75	0.32	90.13	0.85	0.00	81.87	0.81	0.03
$RCI_{consecutive} - F3$	83.00	0.78	0.02	81.50	0.80	0.24	85.96	0.80	0.00	79.85	0.78	0.12
$RCI_{consecutive} - F4$	93.74	0.91	0.00	90.26	0.89	0.00	94.23	0.91	0.00	91.16	0.90	0.00
<i>Only mobility and outings</i>												

Vars	AdaBoost			kNN			LinearSVM			MLP		
	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.	Acc.	F _{score}	Sens.
<i>RCI_{baseline} – total</i>	71.90	0.64	0.09	71.06	0.72	0.59 *	73.59	0.62	0.00	70.82	0.69	0.37
<i>RCI_{baseline} – F1</i>	63.18	0.59	0.22	66.01	0.66	0.51	66.17	0.54	0.02	60.68	0.58	0.31
<i>RCI_{baseline} – F2</i>	93.90	0.91	0.00	93.14	0.93	0.30	94.23	0.91	0.00	90.67	0.90	0.11
<i>RCI_{baseline} – F3</i>	82.08	0.76	0.00	78.35	0.78	0.36	84.29	0.77	0.00	74.19	0.73	0.12
<i>RCI_{baseline} – F4</i>	94.74	0.93	0.00	94.23	0.92	0.00	95.06	0.93	0.00	92.92 *	0.92 *	0.00
<i>RCI_{consecutive} – total</i>	70.69	0.64	0.19	59.63	0.60	0.35	71.92	0.60	0.00	65.42	0.64	0.32
<i>RCI_{consecutive} – F1</i>	66.81	0.60	0.12	64.47	0.64	0.38	70.68	0.59	0.00	67.22	0.65	0.34
<i>RCI_{consecutive} – F2</i>	88.72	0.85	0.00	86.49	0.84	0.11	90.13	0.85	0.00	83.94	0.82	0.03
<i>RCI_{consecutive} – F3</i>	85.80	0.79	0.00	76.29	0.76	0.19	85.96	0.80	0.00	79.51	0.77	0.08
<i>RCI_{consecutive} – F4</i>	94.15	0.91	0.00	91.67	0.90	0.00	94.23	0.91	0.00	92.42	0.91	0.01
Only sleep												
<i>RCI_{baseline} – total</i>	68.96	0.62	0.08	68.49 *	0.65	0.19	73.59	0.62	0.00	64.74	0.60	0.11
<i>RCI_{baseline} – F1</i>	59.63	0.55	0.19 *	66.12	0.64	0.35	66.92	0.54	0.00	58.9	0.57	0.30
<i>RCI_{baseline} – F2</i>	92.58	0.91	0.00	91.51	0.90	0.02	94.23	0.91	0.00	91.68	0.90	0.02
<i>RCI_{baseline} – F3</i>	81.89	0.78	0.11	82.98	0.80	0.22	84.29	0.77	0.00	78.45	0.74	0.03
<i>RCI_{baseline} – F4</i>	94.32	0.92	0.00	94.81	0.93	0.00	95.06	0.93	0.00	94.32	0.92	0.00
<i>RCI_{consecutive} – total</i>	64.46	0.58	0.06	71.19 *	0.66 *	0.24	71.92	0.60	0.00	64.75	0.61	0.20
<i>RCI_{consecutive} – F1</i>	68.51	0.62	0.13	66.74	0.62	0.16	71.09	0.59	0.00	64.13	0.59	0.16
<i>RCI_{consecutive} – F2</i>	88.22	0.84	0.00	86.00	0.83	0.00	90.13	0.85	0.00	88.21	0.84	0.00
<i>RCI_{consecutive} – F3</i>	83.57	0.79	0.04	82.56	0.79	0.12	85.96	0.80	0.00	81.94	0.77	0.00
<i>RCI_{consecutive} – F4</i>	92.74	0.91	0.00	93.33 *	0.91 *	0.00	94.23	0.91	0.00	91.49	0.90	0.00
Only overnight patterns												
<i>RCI_{baseline} – total</i>	70.60	0.66	0.19	64.29	0.63	0.27	73.59	0.62	0.00	63.46	0.61	0.21
<i>RCI_{baseline} – F1</i>	57.24	0.52	0.13	57.22	0.56	0.28	66.84	0.54	0.00	59.01	0.57	0.28
<i>RCI_{baseline} – F2</i>	92.24	0.90	0.00	87.05	0.88	0.00	94.23	0.91	0.00	89.19	0.89	0.09
<i>RCI_{baseline} – F3</i>	81.06	0.77	0.11	78.12	0.77	0.27	84.29	0.77	0.00	77.92	0.76	0.20

Vars	AdaBoost			kNN			LinearSVM			MLP		
	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.
$RCI_{baseline} - F4$	94.74	0.93	0.00	94.31	0.92	0.00	95.06	0.93	0.00	93.49	0.92	0.00
$RCI_{consecutive} - total$	65.58	0.60	0.14	69.69	0.68	0.36	71.42	0.60	0.00	66.58	0.65	0.30
$RCI_{consecutive} - F1$	69.42	0.65	0.25	65.01	0.63	0.27	71.01	0.59	0.00	62.32	0.60	0.24
$RCI_{consecutive} - F2$	87.72	0.84	0.00	79.96	0.80	0.00	90.13	0.85	0.00	82.52	0.81	0.00
$RCI_{consecutive} - F3$	83.59	0.79	0.03	81.15	0.80	0.24	85.96	0.80	0.00	79.12	0.76	0.04
$RCI_{consecutive} - F4$	93.65	0.91	0.00	93.56*	0.91*	0.00	94.23	0.91	0.00	91.74	0.90	0.00

* Statistically significant improvement ($p < 0.05$) in comparison to the corresponding pairwise random algorithm.

Table 10

Results for LOSOCV of Reliable IADL-C change detection using activity-specific features.

Vars	AdaBoost			kNN			LinearSVM			MLP		
	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.
<i>Only cook and eat</i>												
$RCI_{baseline} - total$	70.25	0.63	0.06	47.11	0.50	0.34	73.55	0.62	0.00	61.98	0.59	0.09
$RCI_{baseline} - F1$	57.85	0.51	0.05	62.81	0.60	0.23	66.94	0.54	0.00	64.46	0.61	0.25
$RCI_{baseline} - F2$	90.91	0.90	0.00	89.26	0.89	0.00	94.21	0.91	0.00	90.91	0.90	0.00
$RCI_{baseline} - F3$	84.29	0.77	0.00	56.20	0.62	0.42	84.30	0.77	0.00	76.86	0.73	0.00
$RCI_{baseline} - F4$	95.04	0.93	0.00	94.21	0.92	0.00	95.04	0.93	0.00	95.04*	0.93	0.00
$RCI_{consecutive} - total$	69.42	0.61	0.06	49.59	0.52	0.44	71.90	0.60	0.00	70.25*	0.67	0.26
$RCI_{consecutive} - F1$	71.07	0.64	0.11	63.64	0.60	0.17	71.07	0.59	0.00	65.29	0.62	0.17
$RCI_{consecutive} - F2$	85.95	0.83	0.00	85.12	0.84	0.08	90.08	0.85	0.00	85.12	0.83	0.00
$RCI_{consecutive} - F3$	85.12	0.79	0.00	58.68	0.65	0.53	85.95	0.79	0.00	82.64	0.78	0.00
$RCI_{consecutive} - F4$	94.21	0.91	0.00	91.74	0.90	0.00	94.21	0.91	0.00	94.21*	0.91	0.00
<i>Only mobility</i>												
$RCI_{baseline} - total$	73.55	0.62	0.00	71.07	0.71	0.41	73.55	0.62	0.00	70.25	0.70	0.38
$RCI_{baseline} - F1$	58.68	0.54	0.13	59.50	0.60	0.33	66.94*	0.54	0.00	66.94	0.66	0.40
$RCI_{baseline} - F2$	94.21	0.91	0.00	90.90	0.90	0.00	94.21	0.91	0.00	93.39	0.91	0.00
$RCI_{baseline} - F3$	84.29*	0.77	0.00	79.34	0.77	0.16	84.30	0.77	0.00	75.21	0.73	0.05
$RCI_{baseline} - F4$	95.04	0.93	0.00	93.39	0.92	0.00	95.04	0.93	0.00	95.04*	0.93	0.00
$RCI_{consecutive} - total$	66.94	0.59	0.03	64.46	0.64	0.29	71.49	0.60	0.00	72.73*	0.71	0.38
$RCI_{consecutive} - F1$	67.77	0.63	0.14	66.12	0.65	0.31	71.07	0.59	0.00	61.98	0.57	0.09
$RCI_{consecutive} - F2$	90.08	0.85	0.00	85.12	0.83	0.00	90.08	0.85	0.00	87.60	0.85	0.08
$RCI_{consecutive} - F3$	85.95	0.79	0.00	80.17	0.76	0.00	85.95	0.79	0.00	78.51	0.77	0.06
$RCI_{consecutive} - F4$	94.21	0.91	0.00	91.74	0.90	0.00	94.21	0.91	0.00	93.39	0.91	0.00
<i>Only mobility and outings</i>												

Vars	AdaBoost			kNN			LinearSVM			MLP		
	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.
$RCI_{baseline} - total$	70.24	0.64	0.09	69.42	0.70	0.47	73.55	0.62	0.00	66.94	0.66	0.31
$RCI_{baseline} - F1$	57.02	0.55	0.23	64.46	0.65	0.48	66.12	0.53	0.00	65.28*	0.64	0.40
$RCI_{baseline} - F2$	94.21	0.91	0.00	94.21*	0.93	0.14	94.21	0.91	0.00	89.26	0.89	0.00
$RCI_{baseline} - F3$	84.30	0.77	0.00	77.69	0.77	0.26	84.30	0.77	0.00	73.55	0.72	0.05
$RCI_{baseline} - F4$	94.21	0.92	0.00	93.39	0.92	0.00	95.04	0.93	0.00	0.92	91.73	0.00
$RCI_{consecutive} - total$	71.07	0.68	0.26*	59.5	0.60	0.29	71.90	0.60	0.00	68.59	0.67	0.29
$RCI_{consecutive} - F1$	66.94	0.62	0.14	62.8	0.63	0.34	71.08	0.59	0.00	68.59*	0.67	0.31
$RCI_{consecutive} - F2$	90.08	0.85	0.00	89.25*	0.86	0.08	90.08	0.85	0.00	85.12	0.83	0.00
$RCI_{consecutive} - F3$	85.95	0.79	0.00	77.69	0.76	0.06	85.95	0.79	0.00	80.99	0.79	0.12
$RCI_{consecutive} - F4$	94.21	0.91	0.00	90.91	0.90	0.00	94.21	0.91	0.00	90.91	0.90	0.00
Only sleep												
$RCI_{baseline} - total$	67.76	0.59	0.00	66.12*	0.61	0.09	73.55	0.62	0.00	63.63	0.59	0.06
$RCI_{baseline} - F1$	57.85	0.53*	0.13*	61.98	0.59	0.25	66.94	0.54	0.00	51.24	0.51	0.23
$RCI_{baseline} - F2$	91.73	0.90	0.00	94.21*	0.93	0.14	94.21	0.91	0.00	91.74	0.90	0.00
$RCI_{baseline} - F3$	80.17	0.77	0.11	80.99	0.75	0.00	84.30	0.77	0.00	79.34	0.74	0.00
$RCI_{baseline} - F4$	94.21	0.92	0.00	95.04	0.93	0.00	95.04	0.93	0.00	95.04	0.93	0.00
$RCI_{consecutive} - total$	66.12	0.57	0.00	70.25*	0.64	0.12	71.90	0.60	0.00	66.12	0.62	0.15
$RCI_{consecutive} - F1$	67.77	0.63	0.14	67.77	0.62	0.11	71.07	0.59	0.00	66.94	0.63	0.17
$RCI_{consecutive} - F2$	89.26	0.85	0.00	88.43*	0.86	0.08	90.08	0.85	0.00	90.08*	0.85	0.00
$RCI_{consecutive} - F3$	84.3	0.79	0.00	83.47	0.78	0.00	85.95	0.79	0.00	85.12	0.79	0.00
$RCI_{consecutive} - F4$	94.21	0.91	0.00	93.39	0.91	0.00	94.21	0.91	0.00	92.56	0.92	0.14
Only overnight patterns												
$RCI_{baseline} - total$	68.60	0.65*	0.19	61.98	0.60	0.16	73.55	0.62	0.00	61.98	0.57	0.03
$RCI_{baseline} - F1$	61.98	0.58	0.20	55.37	0.53	0.18	66.94	0.54	0.00	47.93	0.47	0.15
$RCI_{baseline} - F2$	90.91	0.90	0.00	88.43	0.88	0.00	94.21	0.91	0.00	89.26	0.89	0.00

Vars	AdaBoost			kNN			LinearSVM			MLP		
	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.	Acc.	F-sc.	Sens.
$RCI_{baseline} - F3$	80.99	0.78	0.16	76.03	0.74	0.11	84.30	0.77	0.00	80.17	0.78	0.16
$RCI_{baseline} - F4$	95.04	0.93	0.00	94.21	0.90	0.00	95.04	0.93	0.00	92.56	0.91	0.00
$RCI_{consecutive} - total$	66.12	0.62*	0.15*	61.16	0.59	0.15	71.07	0.60	0.00	61.16	0.60	0.21
$RCI_{consecutive} - F1$	71.90*	0.68*	0.26*	64.46	0.62	0.20	71.07	0.60	0.00	64.46	0.62	0.20
$RCI_{consecutive} - F2$	88.43	0.85	0.00	80.17	0.80	0.00	90.08	0.85	0.00	85.12	0.83	0.00
$RCI_{consecutive} - F3$	82.64	0.79	0.06	79.34	0.78	0.12	85.95	0.79	0.00	81.82	0.77	0.00
$RCI_{consecutive} - F4$	93.39	0.91	0.00	93.39	0.91	0.00	94.21	0.91	0.00	90.91	0.89	0.00

* Statistically significant improvement (p<0.05) in comparison to the corresponding pairwise random algorithm.

Table 11

Results for 10-fold CV classification of the Reliable IADL-C change detection using a one-class Linear SVM algorithm.

	All features		Only cook and eat		Only mobility	
	Acc.	F _{score}	Acc.	F _{score}	Acc.	F _{score}
<i>RCI_{baseline} - total</i>	34.71	0.54	33.06	0.52	57.03*	0.63*
<i>RCI_{baseline} - F1</i>	31.41	0.56	33.06	0.55	37.19	0.61
<i>RCI_{baseline} - F2</i>	36.36	0.14	40.50	0.19	66.12	0.08
<i>RCI_{baseline} - F3</i>	36.36	0.41	31.41	0.39	50.41*	0.48*
<i>RCI_{baseline} - F4</i>	80.99	0.40	70.25	0.32	82.65	0.42
<i>RCI_{consecutive} - total</i>	28.93	0.51	31.41	0.50	36.36	0.57
<i>RCI_{consecutive} - F1</i>	27.27	0.54	28.93	0.50	28.93	0.56
<i>RCI_{consecutive} - F2</i>	31.41	0.22	26.45	0.24	27.87	0.30
<i>RCI_{consecutive} - F3</i>	28.93	0.35	39.67	0.36	41.32	0.47
<i>RCI_{consecutive} - F4</i>	77.69	0.35	58.68	0.25	79.34	0.36
	Only mobility and outings		Only sleep		Only overnight patterns	
<i>RCI_{baseline} - total</i>	40.50	0.58	32.23	0.56	33.06	0.55
<i>RCI_{baseline} - F1</i>	32.23	0.55	31.41	0.57	33.06	0.57
<i>RCI_{baseline} - F2</i>	80.17*	0.10*	76.86	0.27	26.45	0.13
<i>RCI_{baseline} - F3</i>	52.90*	0.47*	19.84	0.40	23.97	0.35
<i>RCI_{baseline} - F4</i>	68.60	0.31	83.47	0.43	83.47	0.43
<i>RCI_{consecutive} - total</i>	33.88	0.53	31.41	0.55	28.93	0.50
<i>RCI_{consecutive} - F1</i>	33.88	0.57	28.93	0.54	26.45	0.52
<i>RCI_{consecutive} - F2</i>	31.41	0.25	26.45	0.27	28.10	0.24
<i>RCI_{consecutive} - F3</i>	37.00	0.45	18.18	0.37	21.49	0.36
<i>RCI_{consecutive} - F4</i>	66.12	0.28	74.00	0.33	73.55	0.32

* Statistically significant improvement (p<0.05) in comparison to the corresponding pairwise random algorithm.

Table 12

Results for the Reliable IADL-C Change detection for the SMOTE-based oversampled algorithms using all behavioral features

	AdaBoost			kNN			LinearSVM			MLP		
	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.
$RCI_{baseline} - total$	52.07	0.55	0.63	61.00*	0.63*	0.66*	52.89	0.55	0.66	64.46*	0.65*	0.34*
$RCI_{baseline} - F1$	45.45	0.46	0.60	61.16*	0.62*	0.75*	47.11	0.47	0.75	55.37	0.56	0.70
$RCI_{baseline} - F2$	92.56	0.92	0.14	80.99*	0.84*	0.00	89.26	0.90	0.14	86.77	0.88	0.14
$RCI_{baseline} - F3$	65.29*	0.69*	0.32*	58.68*	0.64*	0.58*	54.54	0.61	0.47	55.37*	0.60*	0.05*
$RCI_{baseline} - F4$	83.00	0.88	0.67	94.22*	0.92*	0.00	95.04*	0.93*	0.00	80.17	0.85	0.67
$RCI_{consecutive} - total$	54.55	0.55	0.88	56.00*	0.58*	0.50*	54.54	0.56	0.38	52.07	0.54	0.35
$RCI_{consecutive} - F1$	53.71	0.55	0.34	58.68*	0.61*	0.57*	48.76	0.51	0.40	57.02	0.57	0.29
$RCI_{consecutive} - F2$	77.69	0.81	0.33	67.77*	0.74*	0.25*	86.77	0.84	0.00	85.95*	0.84*	0.08*
$RCI_{consecutive} - F3$	53.72	0.61	0.35	54.00*	0.61*	0.24*	62.00	0.66	0.12	54.55	0.61	0.29
$RCI_{consecutive} - F4$	82.64	0.87	0.57	94.21*	0.91*	0.00	94.21*	0.91*	0.00	81.82	0.86	0.86

* Statistically significant improvement ($p < 0.05$) in comparison to the corresponding pairwise random algorithm.

Table 13

Positive/negative IADL-C change detection results.

	AdaBoost			kNN			LinearSYM			MLP			C4.5		
	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.	Acc.	F_{score}	Sens.
IADL-C	58.70	0.55	0.27	54.01	0.54	0.60*	60.90	0.51	0.05	56.92	0.57	0.52	59.03	0.56	0.33
IADL-C F1	60.73	0.59	0.38	57.33	0.57	0.42	56.20	0.48	0.07	55.80	0.55	0.47	53.44	0.50	0.19
IADL-C F2	65.33	0.62	0.17	57.58	0.59	0.48	71.03	0.62	0.04	60.57	0.59	0.27	64.57	0.61	0.22
IADL-C F3	68.47	0.63	0.12	63.90	0.62	0.26	71.74	0.62	0.01	65.95	0.64	0.28	62.94	0.59	0.10
IADL-C F4	85.52	0.82	0.10	84.89	0.84	0.29	88.07	0.83	0.00	80.54	0.81	0.24	87.92*	0.84*	0.11

* Statistically significant improvement ($p < 0.05$) in comparison to the corresponding pairwise random algorithm.