



HHS Public Access

Author manuscript

IEEE J Biomed Health Inform. Author manuscript; available in PMC 2017 July 01.

Published in final edited form as:

IEEE J Biomed Health Inform. 2016 July ; 20(4): 1188–1194. doi:10.1109/JBHI.2015.2445754.

Automated Clinical Assessment from Smart home-based Behavior Data

Prafulla Nath Dawadi [Student Member, IEEE],

School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA, 99164

Diane Joyce Cook [Fellow, IEEE], and

School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA, 99164

Maureen Schmitter-Edgecombe

Department of Psychology, Washington State University, Pullman, WA, 99164

Prafulla Nath Dawadi: pdawadi@eecs.wsu.edu; Diane Joyce Cook: cook@eecs.wsu.edu; Maureen Schmitter-Edgecombe: schmittere@wsu.edu

Abstract

Smart home technologies offer potential benefits for assisting clinicians by automating health monitoring and well-being assessment. In this paper, we examine the actual benefits of smart home-based analysis by monitoring daily behaviour in the home and predicting standard clinical assessment scores of the residents. To accomplish this goal, we propose a Clinical Assessment using Activity Behavior (CAAB) approach to model a smart home resident's daily behavior and predict the corresponding standard clinical assessment scores. CAAB uses statistical features that describe characteristics of a resident's daily activity performance to train machine learning algorithms that predict the clinical assessment scores. We evaluate the performance of CAAB utilizing smart home sensor data collected from 18 smart homes over two years using prediction and classification-based experiments. In the prediction-based experiments, we obtain a statistically significant correlation ($r = 0.72$) between CAAB-predicted and clinician-provided cognitive assessment scores and a statistically significant correlation ($r = 0.45$) between CAAB-predicted and clinician-provided mobility scores. Similarly, for the classification-based experiments, we find CAAB has a classification accuracy of 72% while classifying cognitive assessment scores and 76% while classifying mobility scores. These prediction and classification results suggest that it is feasible to predict standard clinical scores using smart home sensor data and learning-based data analysis.

Index Terms

Smart home; Machine Learning; Activity Performance; Activities of Daily Living; Automated Clinical Assessment

I. Introduction

Smart home sensor systems provide the capability to automatically collect information about a resident's everyday behavior without imposing any restrictions on their routines. Researchers have designed algorithms that use such collected information to recognize current activities, prompt individuals to perform needed activities, or perform home automation. Another important use of such sensor data is to predict clinical assessment scores or monitor the health of an individual by monitoring the resident's daily behavior or Activities of Daily Living (ADL).

Several clinical studies support a relationship between daily behavior and cognitive and physical health [1]. Everyday activities like cooking and eating are essential ADLs that are required to maintain independence and quality of life. Decline in the ability to independently perform ADLs has been associated with placement in long-term care facilities, shorter time to conversion to dementia, and a lower quality of life for both the functionally-impaired individuals and their caregivers [2].

In this paper, we investigate whether smart home-based behavior data can be used to predict an individual's standard clinical assessment scores. We hypothesize that a relationship does exist between a person's cognitive/physical health and their daily behavior as monitored by a smart home. We monitor the daily behavior of a resident using smart home sensors and quantify their cognitive/physical health status using standard clinical assessments. To validate this hypothesis, we develop an approach to predict the cognitive and physical health assessment scores by making use of real-world smart home sensor data.

We propose a Clinical Assessment using Activity Behavior (CAAB) approach to predict the cognitive and mobility scores of smart home residents by monitoring a set of basic and instrumental activities of daily living. CAAB first processes the activity-labeled sensor dataset to extract activity performance features. CAAB then extracts statistical activity features from the activity performance features to train machine learning algorithms that predict the cognitive and mobility scores. To evaluate the performance of CAAB, we utilize sensor data collected from 18 real-world smart homes with older adult residents. An activity recognition (AR) algorithm labels collected raw sensor data with the corresponding activities.

CAAB utilizes sensor data collected from actual smart homes without altering the resident's routine and environment. Therefore, the algorithmic approach offers an ecologically valid method to characterize the ADL parameters and assess the cognitive and physical health of a smart home resident [3]. To the best of our knowledge, our work represents one of the first reported efforts to utilize automatically-recognized ADL parameters from real-world smart home data to predict the cognitive and physical health assessment scores of a smart home resident.

II. Related Work

The relationship between in-home sensor-based measurements of everyday abilities and corresponding clinical measurements has been explored using statistical tools and

visualization techniques. Researchers have correlated sensor measurements of sleep patterns, gait, and mobility with standard clinical measurements and self-report data. In one such work, Paavilainen et al. [4] monitored the circadian rhythm of activities of older adults living in nursing homes using the IST Vivago WristCare system. In this study, they compared the changes in activity rhythms with clinical observations of subject health status. In a separate study, these researchers [5] studied the relationship between changes in the sleep pattern of demented and non-demented individuals over a 10-day period.

Several other researchers have considered the relationship between sensor-based activity performance and clinical health assessment. For example, Robben et al. [6] studied the relationship between different high-level features representing the location and transition patterns of an individual's indoor mobility behavior with the Assessment of Motor and Process Skills (AMPS) scores. Similarly, Suzuki and Murase [7] compared indoor activities and outings with Mini-Mental State Examination (MMSE) scores. Dodge et al. used latent trajectory modeling techniques to explore the relationship between gait parameters and cognition [8]. Similarly, LeBellego et al. [9] investigated the relationship between indicators such as mobility and agitation with patient health status in a hospital setting.

In other work, researchers such as Galambos et al. [10] developed techniques to visualize long-term monitoring of sensor data including activity level and time spent away from home [10], [11]. Similarly, other researchers have developed techniques to visualize activity and behavioral patterns by monitoring them with smart home sensors [12], [13], and by monitoring consumption of electricity usage [14].

In our earlier work, we demonstrated a correlation between smart home sensor-based performance measures of simple and complex ADLs and validated performance measures derived from direct observation of participants completing the ADLs in a smart home laboratory [15]. Here we extend this prior work by further investigating this relationship between continuous sensor data collected from real-world smart homes and specific components of standard clinical assessment scores.

III. Problem Formulation

We assume that smart home sensors produce a continuous sequence of time-stamped sensor readings, or sensor events. These sensors continuously generate raw sensor events while residents perform their routine activities of daily living. We use an activity recognition algorithm to automatically annotate each of these sensor events with a corresponding activity label. Activity recognition algorithms map a sequence of raw sensor events onto an activity label A_j , where the label is drawn from the predefined set of activities $A = \{A_1, A_2, \dots, A_n\}$. Our activity recognition algorithm generates a label that corresponds to the last event in the sequence (i.e., the label indicates the activity that was performed when the last event was generated). Activities from set A can be recognized even when the resident interweaves them or multiple residents perform activities in parallel.

CAAB extracts activity performance features from activity-labeled smart home sensor data and utilizes these features to predict standard clinical assessment scores. Therefore, there are two steps involved in CAAB:

- Modeling the ADL performance from the activity-labeled smart home sensor data.
- Predicting the cognitive and mobility scores using a learning algorithm.

Activity modeling

We extract a d -dimensional activity performance feature vector $P_i = \langle P_{i,1}, \dots, P_{i,d} \rangle$ to model the daily activity performance of an activity A_i . Observation $P_{i,d,t}$ provides a value for feature d of activity A_i observed on day t ($1 \leq t \leq T$). The set of all observations in P_i is used to model the performance of A_i during an entire data collection period between day 1 and day T .

Additionally, during the same data collection period, standard clinical tests are administered for the resident every m time units, resulting in clinical assessment scores S_1, S_2, \dots, S_p ($p = T/m$). In our setting, the clinical tests are administered biannually ($m = 180$ days). Therefore, the clinical measurements are very sparse as compared to the sensor observations. The baseline clinical measurement, S_1 , is collected after an initial 180 days of smart home monitoring.

Clinical assessment/Clinical assessment scores prediction

CAAB's goal is to accurately predict clinical assessment scores at time k , or S_k , using activity performance data P_i between time points j and k , $j < k$.

CAAB relies on an activity recognition (AR) algorithm to generate labeled data for the performance feature vector that is an integral component of activity modeling. The method for activity recognition is explained briefly later in this paper and explored in detail elsewhere [16]. Here, we utilize our own AR algorithm and focus on the additional steps that comprise CAAB.

IV. Experimental Setup

We use CAAB approach to analyze data collected in our CASAS smart homes¹ [17] and in our corresponding clinical measurements. Below, we explain the smart home test bed, smart home sensor data, and standard clinical data that are collected as a part of the study.

A. CASAS Smart home test bed

The CASAS smart home test beds used in this study are single-resident apartments, each with at least one bedroom, a kitchen, a dining area, and at least one bathroom. The sizes and layouts of these apartments vary between homes. The homes are equipped with combination motion/light sensors on the ceilings and combination door/temperature sensors on cabinets and doors. These sensors in the smart home test beds unobtrusively and continuously

¹<http://casas.wsu.edu>

monitor the daily activities of its residents. The CASAS middleware collects these sensor events and stores the data on a database server. Figure 1 shows a sample layout and sensor placement for one of the smart home test beds.

The residents perform their normal activities in their smart apartments, unobstructed by the smart home instrumentation. Figure 2 provides a sample of the raw sensor events that are collected and stored. Each sensor event is represented by four fields: date, time, sensor identifier, and sensor value. The raw sensor data does not contain activity labels. We use our AR activity recognition algorithm, described in Section V-A, to label individual sensor events with corresponding activity labels.

B. Residents

Residents included 18 community-dwelling seniors (5 females, 13 males) from a retirement community. All participants are 73 years of age or older ($M = 84.71$, $SD = 5.24$, range 73 – 92) and have a mean education level of 17.52 years ($SD = 2.15$, range 12 – 20). At baseline S_1 , participants were classified as either cognitively healthy ($N = 7$), at risk for cognitive difficulties ($N = 6$) or experiencing cognitive difficulties ($N = 5$). One participant in the cognitively compromised group met the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) criteria for dementia [18], while the other four individuals met criteria for mild cognitive impairment (MCI) as outlined by the National Institute on Aging-Alzheimer's Association workgroup [19]. Participants in the risk group had data suggestive of lowered performance on one or more cognitive tests (relative to an estimate of premorbid abilities), along with sensory and/or mobility difficulties.

C. Clinical tests

Clinicians *biannually* administered standardized clinical, cognitive, and motor tests to the residents. The tests included the Timed Up and Go mobility measure (TUG) as well as the Repeatable Battery for the Assessment of Neuropsychological Status measure of cognitive status (RBANS) as detailed in Table II. We create a clinical dataset using TUG and RBANS scores obtained from biannual clinical tests. Figure 3 plots the distribution of these two scores against the ages of the participants.

V. Modeling Activities and Mobility

A. Activity recognition algorithm

Activity recognition algorithms label activities based on readings (or events) that are collected from smart environment sensors. As described earlier, the challenge of activity recognition is to map a sequence of sensor events onto a value from a set of predefined activity labels. These activities may consist of simple ambulatory motion, such as walking and sitting, or complex basic or instrumental activities of daily living, depending upon what type of underlying sensor technologies and learning algorithms are used.

Our activity recognition algorithm, AR [22], recognizes activities of daily living, such as cooking, eating, and sleeping using streaming sensor data from environmental sensors such as motion sensors and door sensors. These motion and door sensors are discrete-event

sensors with binary states (On/Off, Open/Closed). Human annotators label one month of sensor data from each smart home with predefined activity labels to provide the ground truth activity labels for training and evaluating the algorithm. The inter-annotator reliability (Cohen's Kappa) values of the labeled activities in the sensor data ranged from 0.70 to 0.92, which is considered moderate to substantial reliability. We use the trained model to generate activity labels for all of the unlabeled sensor data.

AR identifies activity labels in real time as sensor event sequences are observed. We accomplish this by moving a sliding window over the data and using the sensor events within the window to provide a context for labeling the most recent event in the window. The window size is dynamically calculated based on the current sensor. Each event within the window is weighted based on its time offset and mutual information value relative to the last event in the window. This allows the events to be discarded that are likely due to other activities being performed in an interwoven or parallel manner. We calculate a feature vector using accumulated sensor events in a window from the labeled sensor data collected over a month. The feature vector contains information such as time of the first and last sensor events, temporal span of the window, and influences of all other sensors on the sensor generating the most recent event based on mutual information. Currently, AR recognizes the activities we monitor in this project with 95% accuracy based on 3-fold cross validation. An example of activity-labeled sensor data is presented in Figure 2 [22]. More details on this and other approaches to activity recognition are found in the literature [16].

B. Modeling performances of activities and mobility performances

The first CAAB step is to model the performance of the activities in set A. We model activity performance by extracting relevant features from the activity-labeled sensor data. For each activity $A_j \in A$, we can represent such performance features using the d -dimensional activity performance feature vector $P_j = \langle P_{j,1}, P_{j,2}, \dots, P_{j,d} \rangle$.

Depending upon the nature of the sensor data and the performance window we want to monitor, we can aggregate activity performance P_j for activity A_j over a day, week, or other time period. In our experiments, we aggregate activity performance features over a day period (the time unit is one day). For example, if we calculate the sleep activity performance $P_{i,1,t}$ as the time spent sleeping in the bedroom on day t , the observation $P_{i,1,t+1}$ occurs one day after observation $P_{i,1,t}$. For each individual, we calculate activity performance features for the entire data collection period T for all activities in the activity set A ($1 \leq t \leq T$).

For our experiments, we model activity performance using two ($d=2$) specific activity performance features, a time-based feature and a sensor-based feature $\{P_{i,1}, P_{i,2}\}$. Feature $P_{i,1}$ represents the duration of activity A_j and $P_{i,2}$ represents the number of sensor events generated during activity A_j . We have provided evidence in previous studies that these two features are generalizable to other activities, are easily interpretable, and can model how the residents perform their daily activities [15]. In addition to capturing activity performance, we also represent and monitor a person's overall mobility. Mobility refers to movement generated while performing varied activities (as opposed to representing a single activity of its own) and is therefore represented using two different types of features: the number of

sensor events triggered throughout the home and the total distance that is covered by movement throughout the course of a single day (see Table III).

C. Selection of ADLs

In this study, we model a subset of automatically-labeled resident daily activities. These activities are sleep, bed to toilet (a common type of sleep interruption), cook, eat, relax, and personal hygiene. We also capture and model a resident's total mobility in the home.

1) Sleep—The effects of aging include changes in sleep patterns that may influence cognitive and functional status. For example, individuals over the age of 75 have been found to experience greater fragmentation in nighttime sleep (e.g., [23]), which concurrently causes decreased total sleep time and sleep efficiency. Sleep problems in older adults can affect cognitive abilities [24] and have been associated with decreased functional status and quality of life. Moreover, individuals with dementia often experience significant disruption of the sleep-wake cycle. Thus, the effects of sleep on the health of older adults are important clinical construct that both clinicians and caregivers are interested in understanding [25].

Using AR, we recognize sensor events that correspond to sleep (in the bedroom, as opposed to naps taken outside the bedroom) and bed-to-toilet activities. We then extract the time spent and number of sensor events features that correspond to these two activities. As listed in Table III, four features model a smart home resident's sleep activity. The value for the time-based sleep feature is calculated as the total number of minutes spent in sleep on a particular day and the value for the sensor-based sleep feature is calculated as the number of sensor events that are triggered over the course of one day while the resident slept. Similarly, the time-based bed to toilet feature is calculated as the total number of minutes spent in bed to toilet activity on a particular day. We exclude the sensor-based feature that calculate number of times sensor events are triggered on bed to toilet activity because our data shows that the number of sensor events generated when performing the bed to toilet activity is often very low. Because of the known importance of sleep and its relationship with physical and cognitive health, we conduct a separate analysis of sleep and bed to toilet parameters from the other activities that are analyzed as a group [25], [26].

2) Mobility—Mobility is the ability of an individual to move around their home environment and the community. Mobility impairments limit an individual's ability to maintain independence and quality of life and are common predictors of institutionalization among older adults [27]. Evidence supports a close connection between executive brain function and walking speed [28]. Therefore, we separately model mobility as an everyday behavioral feature. We model the mobility of a smart home resident based on the number of sensor events they trigger and the total distance they cover in a day while in the home (estimated based on known distances between motion sensors placed in the home). As listed in Table III, the value for the distance-based mobility feature is calculated as the total distance covered by a resident in one day (our aggregation time period) while inside the home. Similarly, the value for the sensor-based mobility feature is calculated as the number of sensor events that a resident triggers over the course of one day while moving around in the home.

3) Activities of Daily Living—Basic activities of daily living (e.g., eating, grooming) and the more complex instrumental activities of daily living (IADLs; e.g., cooking, managing finances), are fundamental to independent living. Data indicate that increased difficulties in everyday activity completion (e.g., greater task inefficiencies, longer activity completion times) occur with older age [29], [30]. Clinical studies have also demonstrated that individuals diagnosed with MCI experience greater difficulties (e.g., increased omission errors) completing everyday activities when compared with healthy controls [31], [32]. Therefore, clinicians argue the importance of understanding the course of functional change given the potential implications for developing methods for both prevention and early intervention [30].

In our work, we consider five activities of daily living (in addition to sleep): cook, eat, personal hygiene, leave home, and relax. We note that the “relax” activity represents a combination of watching TV, reading, and napping that typically takes place in a single location other than the bedroom where the resident spends time doing these activities, such as a favorite chair. We focus on these activities because they are activities of daily living that are important for characterizing daily routines and assessing functional independence. For each of these activities, we calculate the total activity duration. Our data shows the number of sensor events generated when performing these activities is often very low. Thus, for these activities, we exclude features that calculate number of times sensor events are triggered. As listed in Table III, we calculate the value for the time-based ADL feature as the total number of minutes spent in an activity on a particular day.

Algorithm 1 CAAB approach

```

1:  Input: Activity performance features
2:  Output: Statistical activity features
3:  Initialize: Feature vector
4:  //  $T_1$  and  $T_2$  are two consecutive clinical testing time points
5:  Given:  $T_1, T_2$ 
6:  Given: skip size = 1
7:  while  $T_1 < (T_2 - W)$  do
8:    for each activity performance feature do:
9:      Place a window of size  $W$  at  $T_1$ .
10:     Remove missing observations and detrend based on the observations that fall into this window.
11:     Calculate the variance, autocorrelation, skewness, kurtosis and change features (Algorithm 2) using the
        observations in the window.
12:     Append these values to the feature vector.
13:      $T_2 = T_1 + \text{skip size}$ 
14:   end foreach
15: end while
16: return average(Feature matrix)

```

D. Activity feature extraction

The second CAAB step is to extract statistical features from the activity performance vector. CAAB extracts features from the time series-based representation of activity performance

and uses these to train a machine-learning algorithm. Namely, we extract four standard time series features and one new change feature. We will refer to these five features as statistical activity features. Table IV lists the complete set of activity features.

1) Statistical activity features—To calculate the first four features, CAAB runs a sliding window (e.g., window size, $W = 30$ days) over each of the activity performance features listed in Table III and calculates variance, autocorrelation, skewness, and kurtosis using the observations from data that falls within the sliding window. The sliding window starts at one clinical assessment time point and ends at the next assessment time point, thus capturing all of the behavior data that occurred between two subsequent assessments. For example, CAAB calculates the variance, autocorrelation, skewness, and kurtosis of the duration feature for each activity based on duration observations that fall inside each W -sized data window. CAAB repeats the process and calculates these four statistical activity features for all other activity performance features for all of the activities in set A.

Before calculating these features, CAAB first removes the time series trend from the sliding window observations in order to remove the effect of non-stationary components (e.g. periodic components) in the time series [33]. For this step, CAAB fits a Gaussian or a linear trend to the data within the sliding window. CAAB then detrends the data by subtracting the fitted trend from the data. CAAB slides the window by one day (skip size=1) and re-computes all of the statistical activity features. For each feature, CAAB slides a window through the sensor home data and computes the final feature values as an average over all of the windows. Algorithm 1 explains the steps.

In addition to these standard four different time series features, we propose a fifth feature, a change-based feature, to characterize the amount of change in an individual's activity performance. Algorithm 2 details the steps in calculating this new feature. In order to compute this feature, CAAB uses a sliding window of size W days and divides an activity performance feature observations that fall in W into two different groups. The first group contains feature observations that fall in the first half of W and second group contains feature observations that fall in the other half. CAAB then compares between these two groups of feature observations using a change detection algorithm. For the current work, we use the Hotelling-T test algorithm [34]. However, we can also apply other change detection algorithms. CAAB then slides the window by one day (skip size = 1) and re-computes the change feature. CAAB calculates the final change value as the average over all windows. Similar to the other four statistical activity features computed in the previous section, CAAB computes the value of the change feature for each of the activity performance features listed in Table III.

Algorithm 2 Calculation of change feature

- 1: Input: Activity performance features
- 2: Initialize: CH = []
- 3: // T_1 and T_2 are two consecutive clinical testing time points
- 4: Given: T_1, T_2
- 5: Given: skip size = 1

```

6:    $W$  = window size
7:   while do  $T_1 < (T_2 - W)$ :
8:     for each activity performance feature do:
9:       Place window of size  $W$  at  $T_1$ .
10:      Remove missing values that fall into this window.
11:      Put first half of  $W$  in the group A and second half in the group B.
12:      //Returns True or False.
13:      change = Hotelling T-test (A,B)
14:      append(CH, change)
15:       $T_1 = T_1 + \text{skip size}$ 
16:    end foreach
17:  end while
18:  return average(CH)

```

We note that the change feature is different from the variance feature that CAAB calculates earlier. While variance measures the variability of samples around its mean, the change feature empirically calculates the “chance” of observing a change when two sample groups each of size n from the given activity performance features are compared with each other. Here, a higher amount of detected change indicates a greater chance of detecting changes in the activity performance feature.

E. Clinical assessment

In the final step, CAAB predicts the clinical assessment scores of the smart home residents using the activity performance features computed from the activity labeled sensor data. CAAB first aligns the sensor-based data collection date with the clinical assessment-based data collection date before extracting statistical activity features. After extracting features and aligning the data, CAAB then trains a supervised machine learning algorithm and predicts the clinical assessment scores.

To accomplish this goal, CAAB extracts statistical activity features from the activity performance features that lie between any given two consecutive clinical testing points, t_1 and t_2 . Similarly, it obtains the clinical score S_2 (or S_1) at time point t_2 (or t_1). We consider the pair, statistical activity features and clinical score S_2 , as a point in the dataset and repeat the process for all of the smart home residents and for every pair of the consecutive clinical testing points. Algorithm 3 summarizes the steps involved to prepare the dataset.

Algorithm 3 Training set creation

```

1:   Output: Training set to train the learning algorithm
2:   Input: Activity performance features for all residents
3:   Initialize: Empty training set TrSet
4:   for each resident do
5:     for each consecutive clinical testing point  $T_1$  and  $T_2$  do
6:        $F$  = CAAB (activity performance features between  $T_1$  and  $T_2$ )
7:        $S$  = clinical score( $T_1, T_2$ )
8:       Append( $F, S, \text{TrSet}$ )

```

```

9:     end foreach
10:  end foreach

```

The final step in the CAAB is to predict the clinical assessment scores. CAAB trains a learning algorithm to learn a relationship between statistical activity features and the clinical assessment scores using the dataset that is constructed. In this step, for each resident, at each time point (except the first one), CAAB predicts the clinical assessment scores using a learning algorithm.

We note that CAAB predicts clinical assessment scores based on the relationship that the learning algorithm models between the clinical assessment scores and behavior features. We followed this approach because there are very few clinical observations for a resident. Furthermore, we note that CAAB computes activity performance features by temporally following an individual over a period and computes statistical activity features by comparing past observations with current observations. In this way, CAAB uses an individual as their own baseline for predictive assessment.

VI. Experimental Evaluation

A. Dataset

As explained in Section IV-A, the CASAS middleware collects sensor data while monitoring the daily behavior of 18 smart home senior residents for approximately 2 years. We use the AR activity recognition algorithm to automatically label the sensor events with the corresponding activity labels. By running CAAB on the (activity-labeled) sensor data, we compute activity performance features and extract activity features from them. CAAB then creates a training set by combining the activity features and the corresponding clinical assessment scores (RBANS and TUG) to train a learning algorithm.

B. Prediction

We perform the following four different prediction-based experiments to evaluate the performance of CAAB approach and its components: 1) We first evaluate the overall CAAB performance in predicting clinical assessment scores. Here, we train CAAB using the complete set of available features. We compare results from several representative supervised learning algorithms. 2) We then investigate the importance of different activity feature subsets by observing the resulting performance of CAAB in predicting the clinical assessment scores. 3) Next, we investigate the influence of parameter choices on performance by varying CAAB parameter values and analyzing the impact on prediction performance. 4) In the final experiment, we compare CAAB performance utilizing AR-labeled activities with a baseline method that utilizes random activity labels.

We evaluate all of the above experiments using linear correlation coefficient (r) and mean squared error (RMSE). All performance values are generated using leave-one-out cross validation. The data for each participant is used for training or held out for testing, but is not used for both to avoid biasing the model. We use the following methods to compute our performance measures.

- Correlation coefficient(r): The correlation coefficient between two continuous variables X and Y is given as: $r_{X,Y} = \frac{cov(X,Y)}{\sigma_x \sigma_y}$ where σ_x and σ_y are the standard deviations of X and Y and $cov(X,Y)$ is the covariance between X and Y . In our experiments, we evaluate the correlation between the learned behavior model and clinical assessment scores. We will interpret the experimental results based on the absolute value of the correlation coefficient because our learning algorithm finds a nonlinear relationship between statistical activity features and the clinical assessment scores.
- Root Mean Squared Error (RMSE): If \hat{y} is a size- n vector of predictions and y is the vector of true values, the RMSE of the predictor is

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

1) Overall CAAB prediction performance—To validate the overall performance of CAAB performance, we compute correlations between the CAAB-predicted clinical assessment scores and the provided clinical assessment scores using the complete set of activity features and three different supervised learning algorithms:

- Support Vector Regression (SVR): Support vector regression uses support vector machine algorithm to make numeric predictions. The learning model can be expressed in term of support vectors and kernel functions can be used to learn a non-linear function. SVR uses the epsilon insensitive loss function that ignores errors that are smaller than threshold $\epsilon > 0$. We use a linear kernel to generate all our prediction-based performance results [35].
- Linear Regression (LR): Linear regression models the relationship between the class and the features as the weighted linear combination of the features. The weights are calculated from the training data often using the least square approach.
- Random Forest (RF): Random forest builds an ensemble learner by creating multiple decision trees on different bootstrap samples of the dataset. It averages the predictions from these decision trees to make the prediction [35].

As listed in Table V, we observe that the performances of the learning algorithms in predicting the clinical assessment scores are similar. We also observe that the correlation values are all statistically significant. Because SVR performed best overall, we will conduct all of the remaining experiments using this approach. Additionally, we observe that the overall correlation between the predicted TUG scores and the actual TUG scores are weaker than the predicted RBANS and actual RBANS scores. The weaker correlation is likely due to the fact that there are only two activity performance features (mobility and leave home) that represent the mobility of an individual. Other activities such as cook, bed to toilet, and relax do not adequately represent the mobility of a resident.

2) CAAB prediction performance based on activity feature subsets—We

perform a second set of prediction-based experiments using different subsets of statistical activity features to study and find the important sets of features as listed as follows:

1. We evaluate the prediction performances of the learning algorithm when it is trained using different subsets of statistical activity features.
2. We evaluate the result of using statistical activity features that belong to various subsets of ADLs.

In the first experiment, we study the significance of five major types of statistical activity features (autocorrelation, skewness, kurtosis, variance, and change) that CAAB extracts from the activity performance features. To perform this experiment, we create five different training sets, each of which contains a subset of the statistical activity features. For example, the first training set contains all of the variance-based features; the second training set contains all of the autocorrelation-based features etc. Using these training sets, we train five separate support vector machines. As listed in Table VI, we note that the performance of the SVR in predicting clinical assessment scores using the variance of the activity features is strong as compared to other major types of statistical activity features. Therefore, we hypothesize that the variance of activity performance is an important predictor. Additionally, we observe that skewness-based feature is important for predicting TUG clinical scores while it was slightly weaker for RBANS predictions.

For the second CAAB feature-based experiment, we study the relationship between the clinical assessment scores and the statistical activity features subsets that belong to various groups of ADLs. We create nine different ADL groups, each of which contains a combination of one or more activities (out of seven activities) and/or mobility. For each combination, we create a training set containing all statistical activity features belonging to the activities in that combination. In total, we create nine different training sets. As listed in Table VII, we make the following three observations:

1. In terms of single variables, sleep had the highest correlation with RBANS ($r = 0.51$). In contrast, mobility showed little correlation with either clinical score.
2. We observe that correlation is higher when we combine variables. Specifically, including automatically-recognized ADLs improved the correlation further for both RBANS ($r = 0.61$) and TUG ($r = 0.48$). RBANS showed highest correlation when all features are used ($r = 0.72$).
3. In the case of TUG, the only two variable combinations that lacked a significant correlation included mobility. Once again, adding automatically-recognized activities generally increases the correlation.

These results show that a relationship exists between RBANS and TUG clinical assessment scores with combined smart home-based parameters of sleep and ADLs. Our observations are interesting and align with results from prior clinical studies that have found relationships between sleep and ADL performance with cognitive and physical health [24], [36]. Furthermore, we also note that our observations are computed by making use of automated

smart home sensor data and actual clinical assessment scores. The smart home sensor data are ecologically valid because the smart home collects data from the real world environment and CAAB extracts features without governing, changing, or manipulating the individual's daily routines.

3) CAAB performance using different parameters—We perform two different experiments to study the effect of parameter choices on CAAB. In these two experiments, we train the learning algorithm using the complete set of features. We first study how the activity features extracted at different window sizes will affect the final performances of the learning algorithm. Second, we repeat the steps of the first experiment to study the effect of using different trend removal techniques.

In the first experiment, we compare performance using different window sizes and the SVR learning algorithm. We summarize the results in Figure 4. We observe that the strength of the correlation between the actual clinical assessment scores and predicted scores using features derived from smaller and mid-sized window is stronger than the larger-sized windows. One possible explanation is that larger windows encapsulate more behavior trends and day-to-day performance variation may be lost. Therefore, we use mid-sized (30 for RBANS and 55 for TUG) windows for all of our experiments.

In the second experiment, we compare three different trend removal techniques. We create three different training sets that result from removing a Gaussian trend, a linear trend, and no trend removal. The results are showed in Figure 4. We observe that the strength of the correlation coefficients is stronger and often RMSE values are smaller when we remove a Gaussian trend from the observations. Thus, in all of our remaining experiments, we remove a Gaussian trend from the data.

C. CAAB performance using random activity labels

In our final prediction experiment, we compare CAAB performance using AR-labeled activities to CAAB performance using random activity labels. There are three main objectives of this experiment. First, we want to determine the importance of the role that the AR algorithm plays in CAAB. Second, we want to verify that CAAB is not making predictions based on random chance. Third, we let prediction performance based on random activity labels serve as a base-line or lower bound performance for comparison purposes. We expect CAAB performance using AR-labeled activities to significantly outperform the baseline performance.

To perform this experiment, we create a training set in which the statistical activity features (shown in Table III) are calculated from the sensor data that is randomly labeled with the activity instead of using AR algorithm to automatically generate activity labels. We performed this experiment using the following three steps: 1) We label raw sensor events by randomly choosing the activity labels from the activity set. We choose an activity assuming a uniform probability distribution over all activity classes. 2) We extract statistical activity features from the sensor data labeled with the random activities. 3) We train SVR using the statistical features and use clinical assessment scores as ground truth. Performance measures are computed as described in the previous sections.

As shown in Figure 5, we see that the strength of the correlation coefficients between predicted and actual clinical assessment scores are weak and that the RMSE values are high for the random approach. We also observed that the performances of the learning algorithms trained with features obtained from the AR labeled activities are significantly better than the random labels. Thus, we conclude that activity recognition plays a vital role in CAAB and that the CAAB predictions using statistical activity features extracted from AR labeled sensor data are meaningful and not obtained by chance.

VII. Classification Experiments

To evaluate the performance of CAAB using various classification-based experiments to evaluate, we first discretize the continuous clinical assessment scores into two binary classes and then use a learning algorithm to classify smart home residents into one of these two clinical groups. Performing these experiments allows us to use traditional supervised learning-based methods and performance measures to evaluate CAAB, in contrast with the regression approaches that are utilized earlier in the paper. We train the learning algorithms using the CAAB-extracted statistical activity features. For all of the classification-based experiments, we use a support vector machine (SVM) as the learning algorithm [35]. SVM identify class boundaries that maximize the size of the gap between the boundary and data points. We perform the following four different classification experiments: 1) We first evaluate classification performances of the SVM in classifying discretized RBANS and TUG clinical assessment scores when they are trained with different subsets of statistical activity features and activity performance features. 2) In the second experiment, we repeat the first experiment by discretizing RBANS and TUG scores into binary classes at different thresholds. 3) Next, we study the classification performances of the learning algorithms trained using the activity features obtained from the sensor data labeled with random activities. 4) Finally, we evaluate the classification performance (error) by using a permutation-based test to ensure that the accuracy results are not obtained by a chance.

We evaluate the classification performance of the learning algorithm using area under the curve, G-mean, accuracy and error and generate them using leave-one-out cross-fold validation.

- ROC curves assess the predictive behavior of a learning algorithm independent of error cost and class distribution. The area under the ROC curve (AUC) provides a measure that evaluates the performance of the learning algorithm independent of error cost and class distribution.
- G-Mean is the square root of the product of the true positive and true negative rate [35].

$$\text{G-Mean} = \sqrt{(\text{true positive rate} \times \text{true negative rate})}$$
- Accuracy is the percent of the correct predictions made by the learning algorithm by the total number of predictions. Accuracy = #Correct predictions/#Total predictions

- Error is the percent of the incorrect predictions made by the learning algorithm by the total number of predictions. $\text{Error} = 1 - \text{Accuracy}$

1) CAAB classification performance based on feature subsets

Similar to the prediction-based experiments, we first study the importance of different subsets of statistical activity features and subsets of activities. For the first experiment, we discretize clinical assessment scores (RBANS and TUG) into binary classes using an equal frequency binning technique. We then train multiple SVMs to learn the relationship between CAAB-extracted activity features and these discretized clinical assessment scores. We make three observations based on the classification performances presented in Tables VIII and IX.

1. From Table IX, we observe that the performance of the learning algorithm that is trained with the AR-labeled activities including sleep and ADLs performs generally better than using other single variables.
2. From Table VIII, we observe that the classification performances of the SVM when trained with variance-based activity features are better for both RBANS and TUG scores. It appears that skewnewss-based feature is only important for classifying RBANS clinical scores and not for the TUG classifications.
3. We note that the CAAB performance in the classification-based experiments involving smart home-based parameters of sleep and ADLs are similar to the performances in the prediction-based experiments.

In the second experiment, we evaluate the impact of CAAB performance of discretizing the continuous clinical assessment scores into binary classes at different cutoff thresholds. The objective of this experiment is to identify the range of thresholds that the learning algorithm can discriminate. For this experiment, we first discretize RBANS and TUG scores into binary classes at different thresholds. For this experiment, we use all the features to train the SVM with AdaBoost and generate performance metrics using leave one out cross validation. We use SVM/AdaBoost to handle the class imbalance in the dataset if there exists one [35]. The AdaBoost algorithm improves the accuracy of the “weak” learner by assigning greater weight to the examples that the learning algorithm initially fails to correctly classify [35]. The advantages of boosting the classifier to learn an imbalanced class is that since boosting weights the samples, it implicitly performs both up-sampling and down-sampling with little information loss and is also known to prevent overfitting [35]. As showed in Figure 6 we observe some variations in the performance of the learning algorithms when they are trained with class labels that were discretized at different thresholds; however, the majority of the classification performances are better than random classification performances (i.e., 50% accuracy for binary classes).

Additionally, based on Figure 6, we make four more observations:

- CAAB performance is generally better when the RBANS clinical score is discretized at thresholds within the lower range of RBANS (85 – 100) performances and within the higher range of RBANS (125 – 130)

performances. It appears that the learning algorithm does successfully distinguish between the two extreme groups.

- CAAB classification performance is best when the continuous TUG clinical score is discretized at scores 12 and 17. We note that a score of 12 and above on the TUG puts individuals into the falls risk category [38]. Given that the TUG test measures the time that is required to comfortably complete the Timed Up and Go task, it appears that the learning algorithm can discriminate between the “slow performers” and the “fast performers.”
- However, we note that similar to the prediction-based experiment, performance of the classifier in classifying TUG based scores is weaker than the performance while classifying RBANS scores. As we mention previously, this weaker performance is likely due to the fact that there are only two activity performance features (mobility and leave home) that represent the mobility of an individual.
- Additionally, we note that CAAB performance in classifying both TUG and RBANS clinical labels are moderate to poor when the clinical scores are discretized into binary classes at the intermediate thresholds. We obtain moderate classification performances because the two classes are more likely to have “similar” activity performance and are therefore harder to distinguish from each other.

In the fourth experiment, we compare classification performance using AR-labeled activities and random activity labels. Similar to the prediction-based experiment, we expect the classification performance based on AR labeled activities to outperform the random method. As illustrated in Figure 7, we observe that AR-based classification outperforms classification with random activity labels and that the results are similar to the earlier regression-based experiments (t-test on g-mean, $p < 0.05$).

2) Permutation-based test

In the final experiment, we determine whether the aforementioned performance results are obtained because of chance, rather than because of the effectiveness of CAAB. With the permutation-based evaluation method, we calculate a p-value to test a null hypothesis about the relationship between the class labels and features. This p-value is calculated as a fraction of times that the performance of CAAB on the dataset that is obtained by shuffling (permuting) the class labels exceeded the performance of CAAB on the original dataset. Similar to the first classification-based experiment, we first discretize RBANS at a threshold of 105.5 and TUG at a threshold of 12.5 using an equal frequency binning technique. We perform a test proposed in Ojala and Garriga [39].

H: We randomly permute the class labels to study the relationship between class labels and the features. The null hypothesis is that there exists no relationship between the data and the class labels.

Table X presents the results from the AR annotated data. Based on the null hypotheses H, we make the following observation: the statistically significant ($p < 0.05$) result for the null

hypothesis (H) indicates that there exists a relationship between the sensor-based activity performance and discretized RBANS and TUG labels.

We repeat this experiment using activity features derived from randomly-labeled activities. Table XI lists the results. Based on the p-values, we fail to reject the null hypothesis (H) that there exists no relationship between the class labels and features. Thus, we conclude that there exists a relationship between the smart home sensors-based activity features and standard clinical assessment scores (RBANS and TUG) and that the performance results are not obtained by chance.

VIII. Conclusions and Future Works

In this paper, we described our CAAB approach to modeling a person's activity behavior based on smart home sensor data. CAAB collects sensor data, models activity performance, extracts relevant statistical features, and utilizes supervised machine learning to predict standard clinical assessment scores. This represents a longitudinal approach in which a person's own routine behavior and changes in behavior are used to evaluate their functional and mobility-based health. We validate our approach by performing several classification and prediction-based experiments. We found statistically significant correlations between CAAB-predicted and clinician-provided RBANS and TUG scores.

Our experiments are conducted using smart home data from 18 smart home residents and the majority of residents are cognitively healthy. Future work will include validation on larger population sizes encompassing a greater period of time. We note that CAAB is not intended to replace existing clinical measurements with the smart home-based predictions but may provide a tool for clinicians to use. We also note that an advantage of CAAB is that sparsely-measured clinical scores can be enhanced using the continuously-collected smart home data and predictions. In the future, we will explore the clinical utility of smart home-based predictions and the role it can play in helping clinicians to make informed decisions.

Acknowledgments

This work was supported in part by grants from the National Institutes of Health (R01EB015853 and R01EB009675) and by a grant from the National Science Foundation (1064628).

References

1. Schmitter-Edgecombe M, Parsey C, Lamb R. Development and psychometric properties of the instrumental activities of daily living: compensation scale. *Archives of clinical neuropsychology: Journal of the National Academy of Neuropsychologists*. Dec; 2014 29(8):776–92.
2. Ouchi Y, Akanuma K, Meguro M, Kasai M, Ishii H, Meguro K. Impaired instrumental activities of daily living affect conversion from mild cognitive impairment to dementia: the Osaki-Tajiri Project. *Psychogeriatrics*. Mar; 2012 12(1):34–42. [PubMed: 22416827]
3. Chaytor N, Schmitter-Edgecombe M, Burr R. Improving the ecological validity of executive functioning assessment. *Archives of clinical neuropsychology*. Apr; 2006 21(3):217–27. [PubMed: 16554143]
4. Paavilainen P, Korhonen I, Lötjönen J, Cluitmans L, Jylhä M, Särelä A, Partinen M. Circadian activity rhythm in demented and non-demented nursing-home residents measured by telemetric actigraphy. *Journal of sleep research*. Mar; 2005 14(1):61–68. [PubMed: 15743335]

5. Paavilainen P, Korhonen I, Partinen M. Telemetric activity monitoring as an indicator of long-term changes in health and well-being of older people. *Gerontechnology*. 2005; 4(2):77–85.
6. Robben, S.; Pol, M.; Kröse, B. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication – UbiComp '14 Adjunct. New York, New York, USA: ACM Press; Sep. 2014 Longitudinal ambient sensor monitoring for functional health assessments; p. 1209-1216.
7. Suzuki T, Murase S. Influence of outdoor activity and indoor activity on cognition decline: use of an infrared sensor to measure activity. *Telemedicine journal and e-health: journal of the American Telemedicine Association*. 2010; 16(6):686–690.
8. Dodge HH, Mattek NC, Austin D, Hayes TL, Kaye JA. Inhome walking speeds and variability trajectories associated with mild cognitive impairment. *Neurology*. Jun; 2012 78(24):1946–1952. [PubMed: 22689734]
9. LeBellego G, Noury N, Virone G, Mousseau M, Demongeot J. A Model for the Measurement of Patient Activity in a Hospital Suite. *IEEE Transactions on Information Technology in Biomedicine*. Jan; 2006 10(1):92–99. [PubMed: 16445254]
10. 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. [PubMed: 24049513]
11. Wang S, Skubic M, Zhu Y. Activity density map visualization and dissimilarity comparison for eldercare monitoring. *IEEE Transactions on Information Technology in Biomedicine*. Jul; 2012 16(4):607–614. [PubMed: 22547460]
12. Chen, C.; Dawadi, P. 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). IEEE; Mar. 2011 CASASviz: Web-based visualization of behavior patterns in smart environments; p. 301-303.
13. Kanis, M.; Robben, S.; Hagen, J.; Bimmerman, A.; Wagelaar, N.; Kröse, B. Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on. Venice, Italy: 2013. Sensor Monitoring in the Home: Giving Voice to Elderly People; p. 97-100.
14. Noury N, Berenguer M, Teyssier H, Bouzid M-J, Giordani M. Building an index of activity of inhabitants from their activity on the residential electrical power line. *IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society*. Sep; 2011 15(5):758–66.
15. Dawadi P, Cook D, Schmitter-Edgecombe M. Automated cognitive health assessment using smart home smart monitoring of complex tasks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2013; 43(6):1302–1313.
16. Cook, DJ.; Krishnan, NC. Activity Learning: Discovering, Recognizing, and Predicting Human Behavior from Sensor Data. New York: Wiley; 2015.
17. Cook DJ, Crandall AS, Thomas BL, Krishnan NC. CASAS: a Smart Home in a Box. *Computer*. Jul; 2013 46(7):62–69.
18. Association, AP. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. 4th. Vol. 4. Washington, DC: American Psychiatric Association; 2000. ser Diagnostic and statistical manual of mental disorders
19. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association*. May; 2011 7(3):270–9.
20. Randolph, C. Repeatable Battery for the Assessment of Neuropsychological Status Update. San Antonio, Texas: Psychological Corporation; 1998.
21. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*. 1991; 39(2):142–148. [PubMed: 1991946]
22. Krishnan NC, Cook DJ. Activity Recognition on Streaming Sensor Data. *Pervasive and mobile computing*. Feb.2014 10:138–154. [PubMed: 24729780]

23. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*. Nov; 2004 27(7):1255–1273. [PubMed: 15586779]
24. Jelicic M, Bosma H, Ponds RWHM, Van Boxtel MPJ, Houx PJ, Jolles J. Subjective sleep problems in later life as predictors of cognitive decline. Report from the Maastricht Ageing Study (MAAS). *International journal of geriatric psychiatry*. Jan; 2002 17(1):73–77. [PubMed: 11802234]
25. Deschenes CL, McCurry SM. Current treatments for sleep disturbances in individuals with dementia. *Current psychiatry reports*. Feb; 2009 11(1):20–26. [PubMed: 19187704]
26. Martin JL, Fiorentino L, Jouldjian S, Josephson KR, Alessi CA. Sleep quality in residents of assisted living facilities: effect on quality of life, functional status, and depression. *Journal of the American Geriatrics Society*. May; 2010 58(5):829–36. [PubMed: 20722819]
27. Hope T, Keene J, Gedling K, Fairburn CG, Jacoby R. Predictors of institutionalization for people with dementia living at home with a carer. *International journal of geriatric psychiatry*. Oct; 1998 13(10):682–690. [PubMed: 9818303]
28. Scherder E, Eggermont L, Swaab D, van Heuvelen M, Kamsma Y, de Greef M, van Wijck R, Mulder T. Gait in ageing and associated dementias; its relationship with cognition. *Neuroscience and biobehavioral reviews*. Jan; 2007 31(4):485–97. [PubMed: 17306372]
29. McAlister C, Schmitter-Edgecombe M. Naturalistic assessment of executive function and everyday multitasking in healthy older adults. *Neuropsychology, development, and cognition Section B, Aging, neuropsychology and cognition*. Jan; 2013 20(6):735–56.
30. 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. [PubMed: 21729400]
31. Farias ST, Mungas D, Reed BR, Harvey D, Cahn-Weiner D, Decarli C. MCI is associated with deficits in everyday functioning. *Alzheimer disease and associated disorders*. 2006; 20(4):217–223. [PubMed: 17132965]
32. Schmitter-Edgecombe M, Parsey CM. Assessment of functional change and cognitive correlates in the progression from healthy cognitive aging to dementia. *Neuropsychology*. Nov; 2014 28(6): 881–893. [PubMed: 24933485]
33. Dakos V, Carpenter SR, Brock WA, Ellison AM, Guttal V, Ives AR, Kéfi S, Livina V, Seekell DA, van Nes EH, Scheffer M. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS one*. Jan.2012 7(7):e41010. [PubMed: 22815897]
34. Hotelling H. The Generalization of Student's Ratio." *The Annals of. Mathematical Statistics*. Aug; 1931 2(3):360–378.
35. Witten, IH.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*. Second. Morgan Kaufmann Publishers Inc; Jun. 2005 Morgan Kaufmann Series in Data Management Systems
36. Pérès K, Chrysostome V, Fabrigoule C, Orgogozo JM, Dartigues JF, Barberger-Gateau P. Restriction in complex activities of daily living in MCI: impact on outcome. *Neurology*. Aug; 2006 67(3):461–466. [PubMed: 16894108]
37. Stopping elderly accidents, deaths & injuries. Center for Disease Control and Prevention; [Online]. Available: http://www.cdc.gov/homeandrecreationalafety/pdf/steady/timed_up_and_go_test.pdf
38. Ojala M, Garriga GC. Permutation Tests for Studying Classifier Performance. *The Journal of Machine Learning Research*. Mar.2010 11:1833–1863.

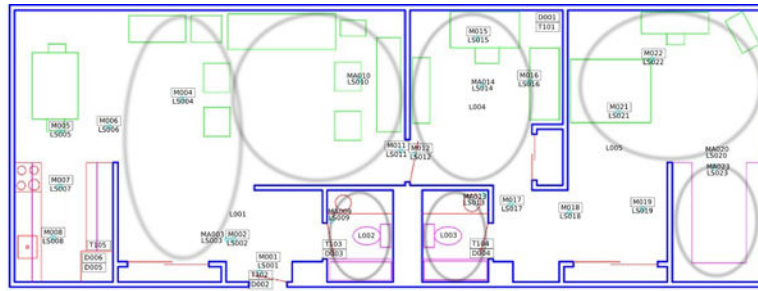


Fig. 1. CASAS smart home floor plan and sensor layout. The location of each sensor is indicated with the corresponding motion (M), light (LS), door (D), or temperature (T) sensor number.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

2009-05-15	09:00:41.05	M042	ON		2009-05-15	09:00:41.05	M042	ON	Cook
2009-05-15	09:00:41.09	M025	OFF		2009-05-15	09:00:41.09	M025	OFF	Eat
2009-05-15	09:00:42.05	M002	ON	→	2009-05-15	09:00:42.05	M002	ON	Relax
2009-05-15	09:00:42.82	M028	ON		2009-05-15	09:00:42.82	M028	ON	Cook
2009-05-15	09:00:43.16	M027	ON		2009-05-15	09:00:42.82	M028	ON	Cook

Fig. 2.

Sample raw (left) and annotated (right) sensor data. Sensors IDs starting with *M* are motion sensors and IDs starting with *D* are door sensors.

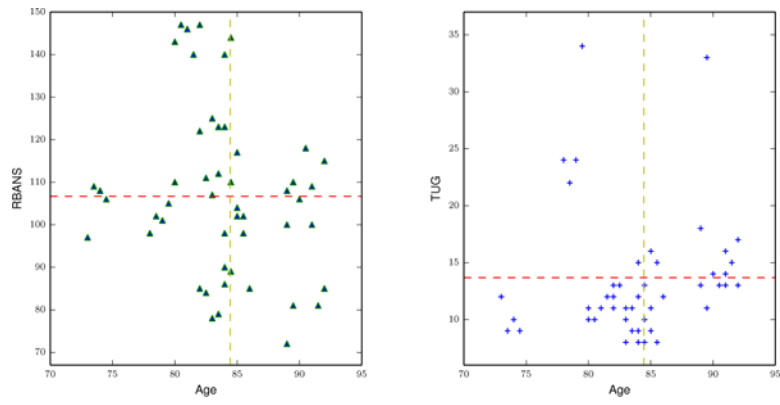


Fig. 3. Distribution of RBANS (left) and TUG (right) clinical assessment scores in the y-axis with respect to age in x-axis. The horizontal line represents a mean clinical score and the vertical line represents the mean age.

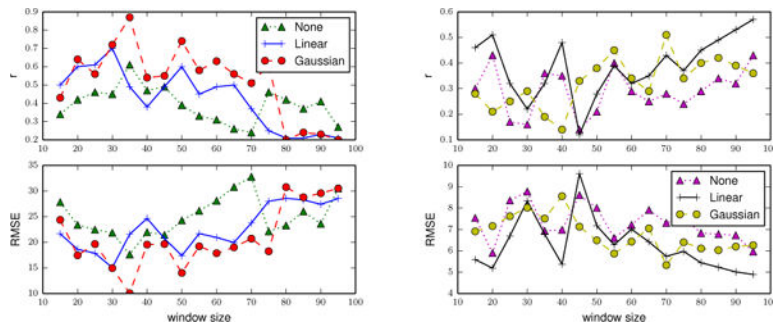


Fig. 4. The correlation coefficients (top) and RMSE (bottom) between predicted and actual RBANS (left) and TUG (right) scores when we use different trend removal techniques and window sizes to train a SVR.

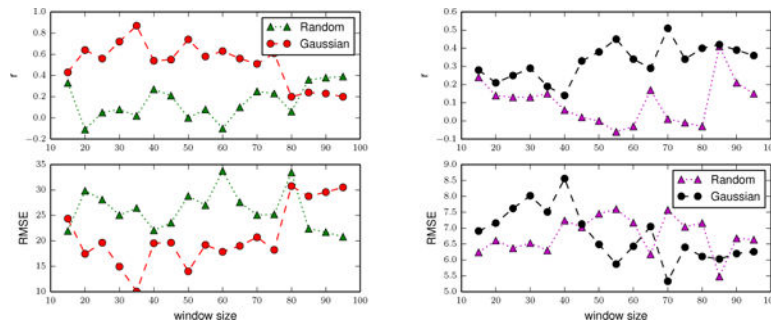


Fig. 5. Correlation coefficients (top) and RMSE (bottom) between SVR-predicted and actual RBANS (left) and TUG (right) scores when we train SVR using features derived from randomly-labeled and AR-labeled activities. We use the complete set of statistical features to train the SVR.

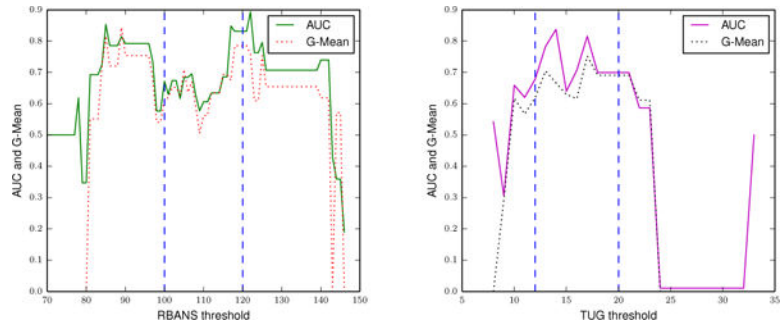


Fig. 6. Classification performance (AUC and G-Mean) of the SVM with boosting in classifying the discretized RBANS (left) and TUG (right) scores. We discretize the RBANS score into two classes at different thresholds and train the SVM using the complete feature set.

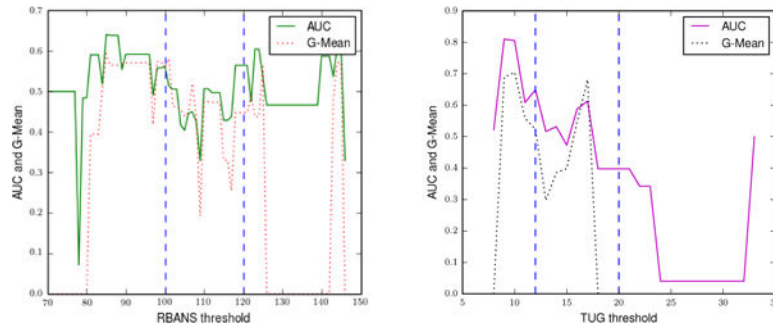


Fig. 7. Classification performance (AUC and G-Mean) of the SVM while classifying RBANS (left) and TUG (right) clinical scores when the SVM is trained using features that are derived from randomly-annotated activities. We use the complete feature set to train the SVMs and discretize the clinical assessment scores into two classes.

TABLE I

Major notations and meanings in CAAB

n	Number of activities
T	Total number of data collection days
A	Set of n activities being modeled
P_i	Activity performance feature vector for activity i modeled over data collection period T
$P_{i,d,t}$	Activity performance feature d for activity i activity on day t
j	Time point at which clinical measurements are made
S_j	Clinical assessment score measured at time point j
W	Sliding window size

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE II

Variables in the standard clinical dataset

Variable name	Description
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)	RBANS [20]. This global measure of cognitive status identifies and characterizes cognitive decline in older adults.
Timed Up and Go (TUG)	TUG [21]. This test measures basic mobility skills. Participants are tasked with rising from a chair, walking 10 feet, turning around, walking back to the chair, and sitting down. The TUG measure represents the time required for participants to complete the task at a comfortable pace.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE III

Activity performance features extracted from the activity-labeled smart home sensor data

Group	Variable	Features
Mobility	Mobility	Total distance traveled, #Total sensor events
Sleep	Sleep	Sleep duration, #Sleep sensor events
	Bed toilet transition	Bed toilet transition duration
	Cook	Cook duration
ADL	Eat	Eat duration
	Relax	Relax duration
	Personal hygiene	Personal hygiene duration
	Leave home	Leave home duration

TABLE IV

Statistical activity features (μ is the mean of the activity performance features p of size n).

Id	Statistical features	Definition	Formula
1	Variance	Variance is the measure of spread.	$\text{Var}(p) = \sum_{k=1}^n (p_i - \mu)^2$
2	Autocorrelation	Autocorrelation(AC) is the similarity between observations that are displaced in time. We calculate autocorrelation at lag 1.	$\text{AC-lag1}(p) = \frac{\sum_{i=1}^{n-1} (p_i - \mu)(p_{i+1} - \mu)}{\sum_{n=1}^n (p_i - \mu)^2}$
3	Skewness	Skewness measures the degree of asymmetry in the distribution of values.	$\text{skewnewess}(p) = \frac{\frac{1}{n} \sum_{i=1}^n (p_i - \mu)^3}{\left(\frac{1}{n} \sum_{i=1}^n (p_i - \mu)^2\right)^{3/2}}$
4	Kurtosis	Kurtosis measures the amount of peakedness of the distribution toward the mean.	$\text{kurtosis}(p) = \frac{\frac{1}{n} \sum_{i=1}^n (p_i - \mu)^4}{\left(\frac{1}{n} \sum_{i=1}^n (p_i - \mu)^2\right)^2}$
5	Change	Change characterizes the amount of change in an individual's activity performance over time.	Algorithm 2

TABLE V

Overall prediction performance of the different learning algorithms

Score Type	Measure	SVR	LR	RF
RBANS	r	0.72 ^{**}	0.64 ^{**}	0.52 ^{**}
	RMSE	14.90	20.25	13.66
TUG	r	0.45 ^{**}	0.41 [*]	0.41 ^{**}
	RMSE	5.87	7.62	5.22

*
 $p < 0.05$.**
 $p < 0.005$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Correlation coefficient (r) and RMSE values between SVR predicted RBANS and TUG scores when SVR Is trained using different types of statistical activity features

TABLE VI

Score Type	Measure	Change	ACF	Skewness	Kurtosis	Variance	All Features
RBANS	r	0.29	0.17	0.30*	0.21	0.49**	0.72**
	RMSE	25.77	21.39	19.90	25.19	17.76	14.94
TUG	r	0.06	0.05	0.43**	0.06	0.31*	0.45*
	RMSE	6.05	6.12	5.23	6.60	5.56	5.87

* $p < 0.05$,

** $p < 0.005$

TABLE VII

Correlation coefficient (r) and RMSE values between SVR-predicted RBANS and TUG scores when the SVR is trained using features from different activities

Score Type	Measure	Sleep	Mobility	ADL	Mobility + Leave Home	ADL + Leave home	Sleep + Mobility	Sleep + ADL	Sleep + ADL + Leave home	Mobility + ADL	All Features
RBANS	r	0.51**	0.08	0.35*	0.18	0.27	0.41*	0.61**	0.57*	0.50**	0.72**
	RMSE	17.53	21.66	20.15	24.49	22.01	19.55	17.51	19.14	19.47	14.94
TUG	r	0.26	0.05	0.35	0.34*	0.43*	0.20	0.48**	0.41	0.13	0.45*
	RMSE	6.19	6.18	5.48	5.48	5.50	6.57	5.55	6.01	6.79	5.87

* $p < 0.05$,

** $p < 0.005$

classification performance (accuracy and AUC) of the SVM in classifying clinical assessment scores (RBANS and TUG) discretized using equal frequency binning. We train SVM using statistical activity features from all activities.

TABLE VIII

Score Type	Measure	Change	ACF	Skewness	Kurtosis	Variance	All features
RBANS	Accuracy	26.92	57.69	73.07	57.69	63.46	71.75
	AUC	0.27	0.58	0.73	0.58	0.63	0.71
TUG	Accuracy	66.00	42.00	46.00	62.00	62.00	76.00
	AUC	0.65	0.39	0.44	0.60	0.62	0.75

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE IX

classification performance (accuracy and AUC) of the SVM in classifying clinical assessment scores (RBANS and TUG) discretized using equal frequency binning. We train SVM using features from different activities

Score Type	Measure	Sleep	Mobility	ADL	Mobility + Leave Home	ADL+Leave home	Sleep +Mobility	Sleep+ADL	Sleep+ADL+ Leave Home	Mobility +ADL	ALL
RBANS	Accuracy	76.92	57.69	46.15	61.53	61.53	75.00	73.08	75.00	48.05	71.15
	AUC	0.76	0.57	0.46	0.62	0.62	0.75	0.73	0.75	0.49	0.71
TUG	Accuracy	78.00	62.00	66.00	52.00	52.94	62.00	76.00	80.00	44.00	76.00
	AUC	0.77	0.61	0.64	0.52	0.50	0.62	0.75	0.79	0.43	0.75

TABLE X

Average error and p-value for our test using support vector machines and activity features extracted from the dataset that is derived from AR-annotated activities

Class label	Original		Test 1	p
	Error	Err (std)		
RBANS	0.27	0.52 (0.11)		0.009**
TUG	0.24	0.42 (0.05)		0.019*

* $p < 0.05$,

** $p < 0.005$

TABLE XI

Average error and p-value for our test using support vector machines and activity features extracted from the dataset that is derived from randomly-labeled activities

	Original		Test1
Class label	Error	Err (std)	p
RBANS	0.57	0.53 (0.07)	0.65
TUG	0.38	0.37 (0.11)	0.48

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript