



Published in final edited form as:

Technol Health Care. 2013 ; 21(4): 323–343. doi:10.3233/THC-130734.

Automated Assessment of Cognitive Health Using Smart Home Technologies

Prafulla N. Dawadi^a, Diane J. Cook^{a,*}, Maureen Schmitter-Edgecombe^b, and Carolyn Parsey^b

Prafulla N. Dawadi: prafulla.dawadi@wsu.edu; Diane J. Cook: djcook@wsu.edu; Maureen Schmitter-Edgecombe: schmitter-e@wsu.edu; Carolyn Parsey: cparsey@wsu.edu

^aSchool of Electrical Engineering and Computer Sciences, Washington State University, Pullman, WA, USA

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

Abstract

BACKGROUND—The goal of this work is to develop intelligent systems to monitor the well being of individuals in their home environments.

OBJECTIVE—This paper introduces a machine learning-based method to automatically predict activity quality in smart homes and automatically assess cognitive health based on activity quality.

METHODS—This paper describes an automated framework to extract set of features from smart home sensors data that reflects the activity performance or ability of an individual to complete an activity which can be input to machine learning algorithms. Output from learning algorithms including principal component analysis, support vector machine, and logistic regression algorithms are used to quantify activity quality for a complex set of smart home activities and predict cognitive health of participants.

RESULTS—Smart home activity data was gathered from volunteer participants (n=263) who performed a complex set of activities in our smart home testbed. We compare our automated activity quality prediction and cognitive health prediction with direct observation scores and health assessment obtained from neuropsychologists. With all samples included, we obtained statistically significant correlation ($r=0.54$) between direct observation scores and predicted activity quality. Similarly, using a support vector machine classifier, we obtained reasonable classification accuracy (area under the ROC curve = 0.80, g-mean = 0.73) in classifying participants into two different cognitive classes, dementia and cognitive healthy.

CONCLUSIONS—The results suggest that it is possible to automatically quantify the task quality of smart home activities and perform limited assessment of the cognitive health of individual if smart home activities are properly chosen and learning algorithms are appropriately trained.

*Corresponding author. Tel.: 509-335-4985; fax: 509-335-3818. cook@eecs.wsu.edu.

Keywords

(methodological) machine learning; smart environments; behavior modeling; (medical) cognitive assessment; MCI; dementia

1. Introduction

The maturing of ubiquitous computing technologies has opened the doors for application of these technologies to areas of critical need. One such area is continuous monitoring of an individual's cognitive and physical health. The possibilities of using smart environments for health monitoring and assistance are perceived as “extraordinary” [1] and are timely given the global aging of the population [2–5].

We hypothesize that cognitive impairment can be evidenced in everyday task performance. We also postulate that differences in task performance can be automatically detected when comparing healthy individuals to those with cognitive impairment using smart home and machine learning technologies. This work investigates approaches for automatically quantifying task performance based on sensor data for identifying correlation between sensor-based assessment and assessment based on direct observation. Clinicians are interested in understanding everyday functioning of individuals to gain insights into difficulties that affect quality of life and to assist individuals in completing daily activities and maintaining independence. Everyday functioning encompasses a range of daily functional abilities that individuals must complete to live competently and independently such as cooking, managing finances, and driving. In addition, deficits and changes in everyday functioning are considered precursors to more serious cognitive problems such as mild cognitive impairment (MCI) and dementia [6, 7].

Mild cognitive impairment is considered a transitional stage between normal cognitive aging and dementia [8], and has been associated with impairments in completing complex everyday activities [9]. To date, much of our understanding of everyday activity completion for individuals with cognitive impairment has come from proxy measures of real-world functioning including self-report, informant-report and laboratory-based simulation measures. Though these methods are commonly used to reflect activity performance in realistic settings, such assessment techniques are widely questioned for their ecological validity [10]. For example, self-report and informant-report are subject to reporter bias [11], while data that is collected via simulation measures in a lab or clinical setting may not capture subtle details of activity performance that occur in a home setting [12,13]. Among these methods, direct observation of the individual to determine everyday functional status has been argued to be the most ecologically valid approach [10,14].

During direct observation, clinicians determine how well individuals perform activities by directly observing task performances. If important steps are skipped, performed out of sequence or performed incorrectly, for example, the activity may be completed inaccurately or inefficiently and this may indicate a cognitive health condition. Such errors may include forgetting to turn off the burner, leaving the refrigerator door open, or taking an unusually long time to complete a relatively simple activity.

However, direct observation methods are conducted in the laboratory and trained clinicians administer them. Often, patients visit the lab to get tested. This is a significant limitation of the direct observation methods. In contrast, smart home sensor systems continuously monitor individuals in their natural environment and thus can provide more ecologically valid feedback [10] about their every day functioning to clinicians or caregiver. Clinicians can use such sensor-derived features that reflect everyday behavior to make informed decisions.

The primary aims of this study are 1) to provide automated task quality scoring from sensor data using machine learning techniques and 2) to automate cognitive health assessment by using machine learning algorithms to classify individuals as cognitively healthy, MCI, or dementia based on the collected sensor data. In this paper, we describe the smart home and machine learning technologies that we designed to perform this task. This approach is evaluated with 263 older adult volunteer participants who performed activities in our smart home testbed. The automated scores are correlated with standardized cognitive measurements obtained through direct observation. Finally, the ability of the machine learning techniques to provide automated cognitive assessment is evaluated using the collected sensor data.

2. Related work

A smart home is an ideal environment for performing automated health monitoring and assessment since no constraints need to be made on the resident's lifestyle. Some existing work has employed smart home data to automate assessment. As an example, Pavel et al. [15] hypothesized that change in mobility patterns are related to change in cognitive ability. They tested this theory by observing changes in mobility as monitored by motion sensors and related these changes to symptoms of cognitive decline. Lee and Dey [16] also designed an embedded sensing system to improve understanding and recognition of changes that are associated with cognitive decline.

The ability to provide automated assessment in a smart home has improved because of the increasing accuracy of automated activity recognition techniques [17,18]. These techniques are able to map a sequence of sensor readings to a label indicating the corresponding activity that is being performed in a smart home. In our experiments we rely upon environmental sensors including infrared motion detectors and magnetic door sensors to gather information about complex activities such as cooking, cleaning and eating [18,20,21]. However, the techniques described in this paper can just as easily make use of other types of sensors such as wearable sensors (i.e., accelerometers) which are commonly used for recognizing ambulatory movements (e.g., walking, running, sitting, climbing, and falling) [22].

More recently, researchers have explored the utility of smart phones, equipped with accelerometers and gyroscopes, to recognize movement and gesture patterns [23]. Some activities, such as washing dishes, taking medicine, and using the phone, are characterized by interacting with unique objects. In response, researchers have explored using RFID tags and shake sensors for tracking these objects and using the data for activity recognition [24,25]. Other researchers have used data from video cameras and microphones as well [20].

While smart environment technologies have been studied extensively for the purposes of activity recognition and context-aware automation, less attention has been directed toward using the technologies to assess the quality of tasks performed in the environment. Some earlier work has measured activity accuracy for simple [26] and strictly sequential [27] tasks. Cook and Schmitter-Edgecombe [27] developed a model to assess the completeness of activities using Markov models. Their model detected certain types of step errors, time lags and missteps. Similarly, Hodges et al. [26] correlated sensors events gathered during a coffee-making task with an individual's neuropsychological score. They found a positive correlation between sensor features and the first principal component of the standard neuropsychological scores. In another effort, Rabbi et al. [13] designed a sensing system to assess mental and physical health using motion and audio data.

3. Background

3.1. The smart home testbed

Data are collected and analyzed using the Washington State University CASAS on-campus smart home test bed, an apartment that contains a living room, a dining area, and a kitchen on the first floor and two bedrooms, an office, and a bathroom on the second floor. The apartment is instrumented with motion sensors on the ceiling, door sensors on cabinets and doors, and item sensors on selected kitchen items. In addition, the test-bed contains temperature sensors in each room, sensors to monitor water and burner use, and a power meter to measure whole-home electricity consumption. Item sensors are placed on a set of items in the apartment to monitor their use.

Sensor events are generated and stored while participants perform activities. Each sensor event is represented by four fields: *date*, *time*, *sensor identifier*, and *sensor message*. CASAS middleware collects sensor events and stores in SQL database. All software runs locally using a small Dream Plug computer. Figure 1 shows the sensor layout in the CASAS smart home testbed. All of the activities for this study were performed in the downstairs of the apartment while an experimenter monitored the participant upstairs via a web camera and remotely communicated to the participant using a microphone and speaker.

3.2. Smart home activities

During the experiment, each participant was introduced to the smart home testbed and guided through a preliminary task in order to familiarize the participant with the layout of the apartment. The participant was then asked to perform a sequence of eight activities. Instructions were given before each activity and no further instructions were given unless the participant explicitly asked for assistance while performing the activity. The eight activities are:

1. *Household Chore: Sweep* the kitchen and dust the dining/living room using supplies from the kitchen closet.
2. *Medication Management: Retrieve medicine* containers and a weekly medicine dispenser. Fill the dispenser with medicine from the containers according to specified directions.

3. *Financial Management*: Complete a *birthday card*, write monetary check, and write an address on the envelope.
4. *General Activity*: Retrieve the specified *DVD* from a stack and watch the news clip contained on the DVD.
5. *Household Chore*: Retrieve the watering can from the closet and water all of the *plants* in the apartment.
6. *Telephone Use/Conversation*: Answer the *phone* and respond to questions about the news clip that was watched previously.
7. *Meal Preparation*: Cook a cup of soup using the kitchen microwave and following the package directions.
8. *Everyday Planning*: Select *clothes* appropriate for an interview from a closet full of clothes.

These activities represent instrumental activities of daily living (IADLs) [28] that can be disrupted in MCI, and are more significantly disrupted in AD. As there is currently no gold standard for measuring IADLs, the IADL activities were chosen by systematically reviewing the literature to identify IADLs that can help discriminate healthy aging from MCI [29,30]. All IADL domains evaluated in this study rely on cognitive processes and are commonly assessed by IADL questionnaires [31,32] and by performance-based measures of everyday competency [33,34]. Successful completion of IADLs requires intact cognitive abilities, such as memory and executive functions. Researchers have shown that declining ability to perform IADLs is related to decline in cognitive abilities [19].

In this study, we examine whether sensor-based behavioral data can correlate with the functional health of an individual. Specifically, we hypothesize that an individual without cognitive difficulties will complete our selected IADLs differently than an individual with cognitive impairment. We further postulate that sensor information can capture these differences in quality of activities of daily living and machine learning algorithms can identify a mapping from sensor-based features to cognitive health classifications.

3.3. Experimental setup

Participants for this study completed a three hour battery of standardized and experimental neuropsychological tests in a laboratory setting, followed approximately one week later by completion of everyday activities in the smart home. The participant pool includes 263 individuals (191 females and 72 males), with 50 participants under 45 years of age (YoungYoung), 34 participants age 45–59 (MiddleAge), 117 participants age 60–74 (YoungOld), and 62 participants age 75+ (OldOld). Of these participants, 16 individuals were diagnosed with dementia, 51 with MCI, and the rest were classified as cognitively healthy. Participants took 4 minutes on average to complete each activity while the testing session for eight activities lasted approximately 1 hour.

As detailed in Table 1, the initial screening procedure for the middle age and older adult participants consisted of a medical interview, the clinical dementia rating (CDR) instrument [35], and the telephone interview of cognitive status (TICS) [36].

Interview, testing and collateral medical information (results of laboratory and brain imaging data when available) were carefully evaluated to determine whether participants met clinical criteria for MCI or dementia. Inclusion criteria for MCI (see Table 1) were consistent with the criteria outlined by the National Institute on Aging-Alzheimer's Association workgroup [37] and the diagnostic criteria defined by Petersen and colleagues [38, 39]. The majority of participants met criteria for amnesic MCI (N = 45), non-amnesic (N = 6), as determined by scores falling at least 1.5 standard deviations below age-matched (and education when available) norms on at least one memory measure (see Table 1). Participants with both single-domain (N = 18) and multi-domain (N = 33) MCI (attention and speeded processing, memory, language, and/or executive functioning) are represented in this sample. Participants in the dementia group met diagnostic and statistical manual of mental disorders (DSM-IV-TR) criteria for dementia [40] and scored 0.5 or higher on the clinical dementia rating instrument. The TICS scores for individuals with dementia ranged from 18 to 29 (M=24, std=3.71).

Before beginning each of the 8 IADL activities in the smart home, participants were familiarized with the apartment layout (e.g., kitchen, dining room, living room) and the location of closets and cupboards. Materials needed to complete the activities were placed in their most natural location. For instance, in the sweeping task a broom was placed in the supply closet and the medication dispenser along with cooking tools were placed in the kitchen cabinet.

As participants completed the activities, two examiners remained upstairs in the apartment, watching the activities through live feed video. As the participant completed the activities, the examiners observed the participant and recorded the actions based on the sequence and accuracy of the steps completed. The experimenters also recorded extraneous participant actions (e.g., searching for items in wrong locations). Experimenter-based direct observation scores were later assigned by two coders who had access to the videos. The coders were blind to diagnostic classification of the older adults. Each activity was coded for six different types of errors: critical omissions, critical substitutions, noncritical omissions, non-critical substitutions, irrelevant actions and inefficient actions. The scoring criteria listed in Table 2 were then used to assign a score to each activity. A correct and complete activity received a lower score, while an incorrect, incomplete, or uninitiated activity received a higher score. The final direct observation score was obtained by summing the individual activity scores and ranged from 8 to 32. Agreement between coders for the overall activity score remained near 95% across each diagnostic group, suggesting good scoring reliability.

Figure 2 shows the distribution of the direct observation scores grouped by participant age and cognitive classification. As participants completed the activities, the examiners recorded the time each subtask began and ended. These timings were later confirmed by watching video of the activity. Using this information, a research team member annotated raw sensor events in the data with the label of the subtask that the individual was performing when the event was triggered. Figure 3 shows a sample of the collected raw and annotated sensor data.

4. Extracting features from smart home sensor data

Based on the objectives of the study, we define features that can be automatically derived from the sensor data, reflect activity performance, and can be fed as input to machine learning algorithms to quantify activity quality and assess cognitive health status. These features capture salient information regarding a participant's ability to perform IADLs. Table 3 summarizes the 35 activity features that our computer program extracts and uses as input to the machine learning algorithms. The last feature, *Health status*, represents the target class label that our machine-learning algorithm will identify based on the feature values.

As Table 3 indicates, we included the age of the participant as a discriminating feature because prior research showing age-related effects on the efficiency and quality of everyday task completion [14, 46].

During task completion, participants sometimes requested help from the experimenter triggering a microphone, and this additional help is noted as a feature value. The experimenters also assigned poorer observed activity quality ratings when the participant took an unusually long time to complete the activity, the participant wandered while trying to remember the next step, the participant explored a space repeatedly (e.g., opened and shut a cabinet door multiple times) as they completed a step, or the participant performed a step in an incorrect manner (e.g., used the wrong tool). The smart features are designed to capture these types of errors. The length of the event is measured in time (duration) and in the length of the sensor sequence that was generated (sequence length).

To monitor activity correctness, the number of unique sensor identifiers that triggered events (sensor count) is captured as well as the number of events triggered by each individual sensor (motion, door, and item sensor counts). Finally, for each activity the smart home software automatically determines the sensor identifiers that are related to the activity, or are most heavily used in the context of the activity, by determining the probability that they will be triggered during the activity. The sensors that have a probability greater than 90% of being triggered (based on sample data) are considered related to the activity the rest are considered unrelated. Therefore, the number of unrelated sensors that are triggered is noted as well as the number of sensor events caused by these unrelated sensors while a participant is performing the activity.

Our feature extraction method does not consider individual "activity steps" while extracting features from the activities. As a result, the features are generalizable to any activity and not fine-tuned to the characteristics of a particular task. This means that the method does not have to be fine-tuned for a particular activity and its steps, but rather will consider features of any activity as a whole. As a result, the technique will be more generalizable to new activities. In addition, it is sometimes difficult to differentiate activity steps from environmental sensors. For example, it is difficult to detect individual steps of the outfit selection activity (moving to the closet, choosing and outfit, and laying out clothes) using only motion sensor data.

The list of features shown in Table 3 is extracted for all eight activities. Our machine learning algorithm receives as input a list of values for each of these 35 sensor-derived features and learns a mapping from the feature values to a target class value (health status). In order to train the algorithm and validate its performance on unseen data, ground truth values are provided for the participants in our study. Ground truth data for a participant is generated from a comprehensive clinical assessment, which includes neuropsychological testing data (described previously), interview with a knowledgeable informant, completion of the clinical dementia rating [35, 47], the telephone interview of cognitive status [36], and a review of medical records. Figure 4 highlights the steps of the automated task assessment.

We observe that participants with cognitive disabilities often leave activities incomplete. Features of incomplete activities as thus denoted as missing. In the final dataset, we only include participants completing 5 or more activities (more than half of the total activities). The final dataset contains 47 (2%) missing instances.

5. Automated task assessment

5.1. Method

The first goal of this work is to use machine learning techniques to provide automated activity quality assessment. Specifically, machine-learning techniques are employed to identify correlation between our automatically derived feature set based on smart home sensor data and the direct observation scores. To learn a mapping from sensor features to activity scores, two different techniques are considered: a supervised learning algorithm using a support vector machine (SVM) [51] and an unsupervised learning algorithm using principal component analysis (PCA) [52]. Support vector machines are supervised learning algorithms that learn a concept from labeled training data. They identify boundaries between classes that maximize the size of the gap between the boundary and the data points. A one vs. one support vector machine paradigm is used which is computationally efficient when learning multiple classes with possible imbalance in the amount of available training data for each class.

For an unsupervised approach, PCA is used to model activities. PCA is a linear dimensionality reduction technique that converts sets of features in a high-dimensional space to linearly uncorrelated variables, called principal components, in a lower dimension such that the first principal component has the largest possible variance, the second principal component has the second largest variance, and so forth. PCA is selected for its widespread effectiveness for a variety of domains. However, other dimensionality reduction techniques could also be employed for this task.

The eight activities used for this study varied dramatically in their ability to be sensed, in their *difficulty*, and in their likelihood to *reflect errors* consistent with cognitive impairment. Therefore, instead of learning a mapping between the entire dataset for an individual and a cumulative score, we build eight different models, each of which learns a mapping between a single activity and the corresponding direct observation score. Because the goal is to perform a direct comparison between these scores and the direct observation scores, and because the final direction observation scores represent a sum of the scores for the

individual activities, the score output from our algorithm is also a sum of the eight individual activity scores generated by the eight different learning models.

5.2. Experimental results

Empirical testing is to evaluate the automated activity scoring and compare scores with those provided by direct observation. The objective of the first experiment is to determine how well an automatically-generated score for a single activity correlates with the direct observation score for the same activity. In the second experiment, similar correlation analyses are performed to compare the automatically-generated combined score for all activities with the sum of the eight direct observation scores. In both cases support vector machines with bootstrap aggregation are used to output a score given the sensor features as input.

In addition to these experiments, a third experiment is performed to compare the automatically-generated combined score with the sum of the eight direct observation scores without using demographic features (age, was help provided) in the feature set. This provides greater insight on the role that only sensor information plays in automating task quality assessment. The bootstrap aggregation improves performance of an ensemble learning algorithm by training the base classifiers on randomly-sampled data from the training set. The learner averages individual numeric predictions to combine the base classifier predictions and generates an output for each data point that corresponds to the highest-probability label.

Table 4 lists the correlation coefficient between automated scores and direct observation scores for individual activities and selected participant groups (cognitive healthy, MCI, and dementia) derived using SVM models. We note that correlation scores are stronger for activities that took more time, required a greater amount of physical movement and triggered more sensors such as *Sweep* as compared to activities such as *Card*. For activities like *Card*, errors in activity completion were more difficult for the sensors to capture. Thus, the correlation scores between automated sensor-based scores and direct observation scores in these activities are low. Similarly, we note that the correlation score also varies based on what groups (cognitively health, MCI, dementia) of participants are included in the training set. In almost all activities, the correlation is relatively strong when the training set contains activity sensor data for all three cognitive groups of participants.

Next a combination of all of the performed activities is considered. Table 5 lists the correlation between a sum of the individual activity scores generated by the eight activity SVM models and the direct observation score. Correlations between two variables are relatively strong when the learning algorithm is trained using data from all three cognitive groups. Differences in correlation strength may be attributed to diversity present in the data. A majority of the cognitive healthy participants complete the eight tasks correctly so the training data from this group contains examples of only “well-performed” activities and thus exhibits less diversity. Learning algorithms tend to generalize poorly when data contains little variation and thus classification performance may degrade.

Table 5 also lists the correlation between a sum of individual activity scores generated by eight SVM models and the direct observation score without considering non-sensor-based features (age, help provided). Both Pearson linear correlation and Spearman rank correlation coefficient are calculated to assess relationship between variables. The correlation coefficients are statistically significant when correlations are derived only from the sensor-based feature set and duration does improve the strength of the sensor-based correlation. We conclude that demographic and experimenter-based features do contribute toward the correlation, but a correlation does exist as well between purely sensor-derived features and the direct observation score.

Tables 6 and 7 list the correlation coefficients between our automated scores and direct observation scores when we utilize PCA to generate the automated scores based on sensor features. Similar to the results in Tables 4 and 5, some activities have much stronger correlations than others and the strength of the correlations varies based on which groups are included in the training set. Furthermore, the correlation scores obtained using PCA are statistically significant but not as strong as those obtained from the SVM models. Given the nature of the activities and given that the dimension of sensor derived features is reduced to a single dimension using a linear dimensionality reduction technique, it is likely that during the process some information is lost that otherwise produces a satisfactory correlation performance between direct observation scores and sensor based features. Note that experiments are not performed which involve only participants from the dementia group because the sample size is small. Table 7 list the correlation coefficients between the automated scores and direct observation scores when PCA is utilized to generate the automated scores based on sensor features excluding non-sensor-based features. As before, we note that there is little difference between the sets of correlation coefficients.

These experiments indicate that it is possible to predict smart home task quality using smart home-based sensors and machine learning algorithms. We observe moderate correlations between direct observation score, which is a task quality score assigned by trained clinical coders, and an automated score generated from sensor features. We also note that the strength of the correlation depends on the diversity and quantity of training data. Finally, we also note that apart from the age of the participants, all of the features that are input to the machine-learning algorithm are automatically generated from smart home sensor events.

6. Automated cognitive health assessment

6.1. Method

The second goal of this work is to perform automated cognitive health classification based on sensor data that is collected while an individual performs all eight activities in the smart home testbed. Here, a machine learning method is designed to map the sensor features to a single class label with three possible values: cognitively healthy (CH), mild cognitive impairment (MCI), or dementia (D).

When sensor data that was gathered for the population is visualized (shown in Figure 5), we see the heterogeneity of the data as well as specific differences in activity performance across the eight selected activities. As a result, we hypothesize that a single classifier would

not be able to effectively learn a mapping from the entire data sequence for an individual to a single label for the individual. This is because individual activities vary in terms of difficulty, duration, and inherent variance.

Machine learning researchers use ensemble methods, in which multiple machine learning models are combined to achieve better classification performance than a single model [53]. Here, eight base classifiers are initially created, one for each of the activities, using both a non-linear classifier (in this case, a support vector machine learning algorithm) and a linear classifier (in this case, a logistic regression classifier).

We observe that there is a class imbalance in our training set for cognitive health prediction: there are only 16 individuals in the Dementia group and 51 in the MCI group, relative to the 196 participants in the cognitively healthy group. We note in advance that such imbalance may adversely affect predictive performance as many classifiers tend to label the points with the majority class label. To address this issue, cost sensitive versions of the machine learning algorithms are used for each of the base classifiers. A cost sensitive classifier assigns misclassification costs separately for individual class labels and reweights the samples during training according to this cost. This allows the classifier to achieve overall strong performance even when the training points are not evenly divided among the alternative classes [54], as is the case with this dataset. A meta-classifier then outputs a label (CH, MCI, or D) based on a vote from the base learners.

6.2 Evaluation metrics

A number of evaluation metrics are utilized to validate the proposed methodology. The first, ROC curves, assess the predictive behavior of a learning algorithm independent of error cost and class distribution. The curve is obtained by plotting false positives vs. true positive at various threshold settings. 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.

In a data set with an imbalanced class distribution, g-mean measures the predictive performance of a learning algorithm for both the positive and a negative classes. It is defined as:

$$gmean = \sqrt{(true\ positive\ rate \times true\ negative\ rate)}$$

where the true positive rate and true negative rate represents the percentage of instances correctly classified to their respective classes. Furthermore, we also report if the prediction performance of a learning algorithm is better than random in both negative and positive classes. The classifier predicts a class better than random if the prediction performance, true positive rate, true negative rate, and the AUC value are all greater than 0.5.

6.3. Experimental results

Several experiments are performed to evaluate our automated cognitive health classifier. For all of the experiments we report performance based on overall area under the ROC curve (AUC) and g-mean scores. Values are generated using leave one out validation.

To better understand the differences between each class, this situation is viewed as a set of binary classification problems in which each class is individually distinguished from another. The first experiment evaluates the ability of the classifier to perform automated health assessment using sensor information from individual activities using the support vector machine and logistic regression classifier algorithms.

Tables 8 and 9 summarize the results. Classification performance for cognitively healthy vs. dementia is better than the other two cases. Similarly, performance for classifying MCI and cognitively healthy using SVM learning algorithm in all eight activities is not better than random prediction. Figure 2 shows that there is overlap between the direct observation scores of the healthy older adults and those diagnosed with MCI. MCI is often considered as a transition stage from cognitively healthy to dementia [35, 38]. It is possible that no reliably distinct differences exist between activity performances, or sensors are not able to capture subtle differences in activity performance between those two cognitive groups. Thus, additional experiments are not performed to distinguish between these two groups.

Similar to the results summarized in Section 5, we see that prediction performance for some activities such as *Sweep*, *Dress*, and *Cook* is better than for other activities such as *DVD* and *Medicine*. As explained previously, some of the activities took longer to complete and triggered more sensor events than others making it easier to identify errors, unrelated sensor events, and taking longer to perform the activity. Thus, differences exist in patterns of activity performance and sensors capture them. Our learning algorithm may be able to quantify these differences to distinguish between the different participant groups.

The second experiment evaluates the ability of the ensemble learner to automate health assessment using information from the combined activity set. The results are summarized in Table 10 and Table 11. The classification performance for classifying dementia and cognitively healthy is better than for classifying MCI vs. dementia. In addition, in Table 8 only a few base classifiers have better than random prediction. For each of these tables, costs are reported that yield the most promising results for the classifier. In a third experiment, only base classifiers that have better than random performance are selected. The results are summarized in Table 12 and 13. As shown in Table 12, for predicting MCI vs. dementia only 4 base classifiers are selected while for predicting dementia vs. cognitively healthy 5 base classifiers are selected. Similarly, for the results summarized in Table 13, only 2 base classifiers are selected for predicting dementia vs. cognitively healthy, while 6 base classifiers are selected for predicting MCI vs. dementia and 2 are selected for predicting dementia vs. cognitively healthy. The classification performance of MCI vs. dementia and dementia vs. cognitively healthy improves as compared to the previous two cases.

These experiments indicate that it is possible to perform limited automated health assessment of individuals based on task performance as detected by smart home sensors.

The feature extraction technique described here along with the learning algorithm design achieves good performance at differentiating the dementia and cognitively healthy groups as compared to the other binary comparisons. This limitation might be due to the current coarse-grained resolution of the environment sensors and the current smart home activity design. It might be possible to improve accuracy using different tasks, additional features, more sensors and those sensors that provide finer resolution such as wearable sensors.

7. Discussion and Conclusions

In this work, we have introduced a method to assist with automated cognitive assessment of an individual by analyzing the individual's performance on IADLs in a smart home. We hypothesize that learning algorithms can identify features that represent task-based difficulties such as errors, confusion, and wandering that an individual with cognitive impairment might commit while performing everyday activities. The experimental results suggest that sensor data collected in a smart home can be used to assess task quality and provide a score that correlates with direct observation scores provided by an experimenter. In addition, the results also suggest that machine-learning techniques can be used to classify the cognitive status of an individual based on task performance as sensed in a smart home.

One must carefully interpret the correlation results that are mentioned here. The correlation (r) between smart home features and direct observation score is statistically significant. The correlation coefficient is squared to obtain the coefficient of determination. A coefficient of determination of 0.29 ($r=0.54$) means that the nearly 30% of the variation in the dependent variable can be explained by the variation in the independent variable. The current results show that this method explains nearly 30% variations in the direct observational scores. Unexplained variation can be attributed to limitations of sensor system infrastructures and algorithms.

We have seen that the predictive performance of a learning algorithm varies based on the activity being monitored and the condition of the individual performing the activities. As expected, the prediction accuracies of complex activities that triggered more sensor events were better than the accuracies for activities that trigger fewer sensor events and required less time to complete. Learning algorithms generalize better when trained from sensor rich data and when they are provided with data from a large segment of the population.

In addition, the prediction performance of the learning algorithm is affected by several factors. A primary factor is the class imbalance in our data set. Another contributing factor is missing values that are introduced in the cases when individuals (almost always individuals in the MCI and Dementia groups) do not attempt some of the activities. Finally, the ground truth values are based on human observation of a limited set of activities and may be prone to error. Based on these observations, we conclude that in a testing situation an experimenter needs to select activities with caution, balancing tradeoff between a difficult activity that results in good prediction performance and one that is easy enough for participants with cognitive impairments to complete.

The current approach uses between-subjects differences in activity performance to perform cognitive health assessment and is based on set of non-obtrusive environmental sensors such

as door sensors, motion sensors, and item sensors. The current work demonstrates that it is possible to automatically quantify the task quality of smart home activities and assist with assessment of the cognitive health of individual with a reasonable accuracy given the proper choice of smart home activities and appropriate training of learning algorithms. However, before sensor-based scores would be acceptable for clinical practice, future studies will be needed to increase the strength of correlations between sensor-based and direct observation scores. Given the variability in functional performance evident across individuals, longitudinally collected sensor-based data, which uses the individual as their own control, will likely increase sensitivity to detecting change in functional status and is an avenue that we are pursuing.

Our long term goal is to develop an automated system that assesses activities of daily living in a home environment by using person's own behavioral characteristics to track their activity performance over time and identify changes in activity performance that correlate with changes in their cognitive health status. This current work is a first step in developing such technique. This work supports the hypothesis that the sensors can assist with automated activity assessment. In future work, activity recognition algorithms can be employed to automatically recognize activities performed in a home setting and repeat this experiment in participants' own homes. In addition, we will analyze longitudinal behavioral and activity data collected by installing environmental sensors in real homes. Activity recognition algorithms will be used to recognize activities and avoid manual annotation. We also plan to develop algorithms based on time series analysis to assess activity quality and correlate detected behavioral changes with changes in cognitive or physical health.

Acknowledgments

The authors thank Selina Akhter, Thomas Crowger, Matt Kovacs, and Sue Nelson for annotating the sensor data. We also thank Alyssa Hulbert, Chad Sanders, and Jennifer Walker for their assistance in coordinating the data collection. This material is supported by the Life Sciences Discovery Fund, by National Science Foundation grants 0852172 and 1064628, and by the National Institutes of Health grant R01EB009675.

References

1. Department of Health. Speech by the Rt Hon Patricia Hewitt MP, Secretary of State for Health. Proceedings of the Long-term Conditions Alliance Annual Conference; 2007;
2. Ernst RL, Hay JW. The US economic and social costs of Alzheimer's disease revisited. *American Journal of Public Health*. 1994; 84:1261–1264. [PubMed: 8059882]
3. World Alzheimer Report[Internet]. Alzheimer's Disease International; [cited 2012 October 8]. Available from: www.alz.co.uk/research/worldreport
4. World population ageing[Internet]. United Nations; [cited 2012 October 8]. Available from: www.un.org/esa/population/publications/worldageing19502050
5. Vincent, G.; Velkoff, V. The next four decades – the older population in the United States: 2010 to 2050. US Census Bureau; 2010.
6. 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]
7. Wadley V, Okonkwo O, Crowe M, Vance D, Elgin J, Ball K, Owsley C. Mild cognitive impairment and everyday function: An investigation of driving performance. *Journal of Geriatric Psychiatry and Neurology*. 2009; 22(2):87–94. [PubMed: 19196629]

8. Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, et al. Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*. 2004; 256:240–246. [PubMed: 15324367]
9. Schmitter-Edgecombe M, Woo E, Greeley D. Characterizing multiple memory deficits and their relation to everyday functioning in individuals with mild cognitive impairment. *Neuropsychology*. 2009; 23:168–177. [PubMed: 19254090]
10. Chaytor N, Schmitter-Edgecombe M, Burr R. Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*. 2006; 21(3):217–227. [PubMed: 16554143]
11. Dassel KB, Schmitt FA. The Impact of Caregiver Executive Skills on Reports of Patient Functioning. *The Gerontological Society of America*. 2008; 48:781–792.
12. Marson, D.; Hebert, K. Functional assessment. In: Attix, D.; Welsh-Bohmer, K., editors. *Geriatric Neuropsychology Assessment and Intervention*. New York: The Guilford Press; 2006. p. 158-189.
13. Rabbi, M.; Ali, S.; Choudhury, T.; Berke, E. Passive and in-situ assessment of mental and physical well-being using mobile sensors. *Proceedings of the International Conference on Ubiquitous Computing*. ACM; New York, USA. 2011; p. 385-394.
14. 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*. 2011; 17(5):853–864. [PubMed: 21729400]
15. Pavel, M.; Adami, A.; Morris, M.; Lundell, J.; Hayes, TL.; Jimison, H.; Kaye, JA. Mobility assessments using event-related responses. *Proceedings of the Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare*; 2006; p. 71-74.
16. Lee, ML.; Dey, AK. Embedded assessment of aging adults: A concept validation with stake holders. *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare*; 2010; IEEE; p. 22-25.
17. Cook DJ. Learning setting-generalized activity models for smart spaces. *IEEE Intelligent Systems*. 2012; 27(1):32–38.
18. Kim E, Helal A, Cook DJ. Human activity recognition and pattern discovery. *IEEE Pervasive Computing*. 2010; 9:48–53. [PubMed: 21258659]
19. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12(3):189–1985. [PubMed: 1202204]
20. Alon J, Athitsos V, Yuan Q, Sclaroff S. A unified framework for gesture recognition and spatiotemporal gesture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008; 31:1685–1699. [PubMed: 19574627]
21. Van Kasteren TLM, Englebienne G, Kröse BJ. An activity monitoring system for elderly care using generative and discriminative models. *Personal and Ubiquitous Computing*. 2010; 14(6): 489–498.
22. Maurer, U.; Smailagic, A.; Siewiorek, D.; Deisher, M. Activity recognition and monitoring using multiple sensors on different body positions. *Proceedings of the IEEE Workshop on Wearable and Implantable Body Sensor Networks*; 2006; Washington DC: IEEE Computer Society; p. 113-116.
23. Kwapisz, JR.; Weiss, GM.; Moore, SA. Activity recognition using cell phone accelerometers. *Proceedings of the International Workshop on Knowledge Discovery from Sensor Data*; 2010; p. 10-18.
24. Palmes P, Pung HK, Gu T, Xue W, Chen S. Object relevance weight pattern mining for activity recognition and segmentation. *Pervasive and Mobile Computing*. 2010; 6:43–57.
25. Philipose M, Fishkin KP, Perkowski M, Patterson DJ, Fox D, Kautz H, Hahnel D. Inferring activities from interactions with objects. *IEEE Pervasive Computing*. 2004; 3:50–57.
26. Hodges, M.; Kirsch, N.; Newman, M.; Pollack, M. Automatic assessment of cognitive impairment through electronic observation of object usage. *Proceedings of the International Conference on Pervasive Computing*; 2010; Berlin, Heidelberg: Springer-Verlag; p. 192-209.

27. Cook DJ, Schmitter-Edgecombe M. Assessing the quality of activities in a smart environment. *Methods of Information in Medicine*. 2009; 48(5):480–485. [PubMed: 19448886]
28. Holsinger T, Deveau J, Boustani M, Williams J. Does this patient have dementia? *Journal of the American Medical Association*. 2007; 297:2391–2404. [PubMed: 17551132]
29. Peres K, Chrysostome V, Fabrigoule C, Orgogozo J, Dartigues J, Barberger-Gateau P. Restriction in complex activities of daily living in MCI. *Neurology*. 2006; 67:461–466. [PubMed: 16894108]
30. Tuokko H, Morris C, Ebert P. Mild cognitive impairment and everyday functioning in older adults. *Neurocase*. 2005; 11:40–47. [PubMed: 15804923]
31. Lawton M, Moss M, Fulcomer M, Kleban M. A research and service oriented multilevel assessment instrument. *Journal of Gerontology*. 1982; 37(1):91–99. [PubMed: 7053405]
32. Resiberg B, Finkel S, Overall J, Schmidt-Gollas N, Kanowski S, Lehfeld H, Hulla F, Sclan SG, Wilms H, Heininger K, Hindmarch I, Stemmler M, Poon L, Kluger A, Cooler C, Berg M, Hugonot-Diner L, Philippe H, Erzigkeit H. The Alzheimer's disease activities of daily living international scale (ASL-IS). *International Psychogeriatrics*. 2001; 13:163–181. [PubMed: 11495392]
33. Diehl M, Marsiske M, Horgas A, Rosenberg A, Saczynski J, Willis S. The revised observed tasks of daily living: A performance-based assessment of everyday problem solving in older adults. *Journal of Applied Gerontology*. 2005; 24:211–230. [PubMed: 18160968]
34. Willis, S.; Marsiske, M. *Manual for the Everyday Problems Test*. University Park: Department of Human Development and Family Studies, Pennsylvania State University; 1993.
35. Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*. 1993; 43:2412–2414. [PubMed: 8232972]
36. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*. 1988; 1:111–117.
37. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, 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 and Dementia. *The Journal of the Alzheimer's Association*. 7:270–279.
38. Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, Winblad B. Current concepts in mild cognitive impairment. *Archives of Neurology*. 2001; 58:1985–1992. [PubMed: 11735772]
39. Peterson RC, Morris JC. Mild cognitive impairment as a clinical entity and treatment target. *Archives of Neurology*. 2005; 62:1160–1163. [PubMed: 16009779]
40. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Psychiatric Press; 2000.
41. Benedict, RHB. *Brief Visuospatial Memory Test – Revised*. Odessa, FL: Psychological Assessment Resources, Inc; 1997.
42. Williams, JM. *Memory Assessment Scales professional manual*. Odessa: Psychological Assessment Resources, Inc; 1991.
43. Delis, DC.; Kaplan, E.; Kramer, JH. *Delis-Kaplan, Executive Function System: Examiner's Manual*. The Psychological Corporation; San Antonio, Texas: 2001.
44. Ivnik RJ, Malec JM, Smith GE, Tangalos EG, Petersen RC. Neuropsychological tests' norms above age 55: COWAT, BNT, MAE token, WRAT-R reading, AMNART, STROOP, TMT, and JLO. *The Clinical Neuropsychologist*. 1996; 10:262–278.
45. Smith, A. *Symbol Digit Modalities Test*. Western Psychological Services; Los Angeles, CA: 1991.
46. Wechsler, D. *Wechsler Adult Intelligence Test. 3*. The Psychological Corporation; New York:
47. Lafortune, G.; Balestat, G. *OECD Health Working Papers, No 26*. OEC Publishing; 2007. Trends in severe disability among elderly people: Assessing the evidence in 12 OECD countries and the future implications.
48. Measso G, Cavarzeran F, Zappalà G, Lebowitz BD. The Mini-Mental State Examination: Normative study of an Italian random sample. *Developmental Neuropsychology*. 1993; 9:77–85.
49. Pavel, M.; Adami, A.; Morris, M.; Lundell, J.; Hayes, TL.; Jimison, H.; Kaye, JA. Mobility assessment using event-related responses. *Proceedings of the Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare*; 2006; 2006. p. 71-74.

50. Hughes CP, Berg L, Danzinger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *British Journal of Psychiatry*. 1982; 140:566–572. [PubMed: 7104545]
51. Smola AJ, Schölkopf B. A tutorial on support vector regression. *Statistics and Computing*. 2004; 14:199–222.
52. Maaten L, Postma E, Herik H. Dimensionality reduction: A comparative review. Tilburg University Technical Report, TiCC-TR 2009-005. 2009; 10:1–35.
53. Džeroski S, Ženko B. Is combining classifiers with stacking better than selecting the best one? *Machine Learning*. 2004; 54(3):255–273.
54. Sun Y, Kamel MS, Wong AKC, Wang Y. Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*. 2007; 40:3358–3378.

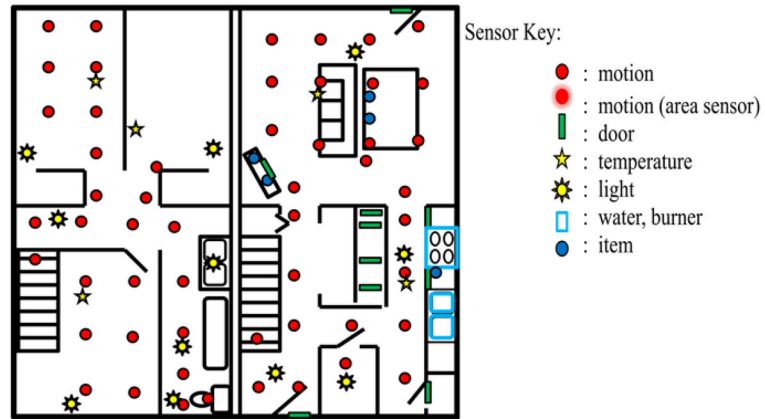


Figure 1. CASAS smart home test-bed floor plan and sensor layout. The test-bed is a two-story apartment with three bedrooms, a bathroom, a living room, dining room, and kitchen.

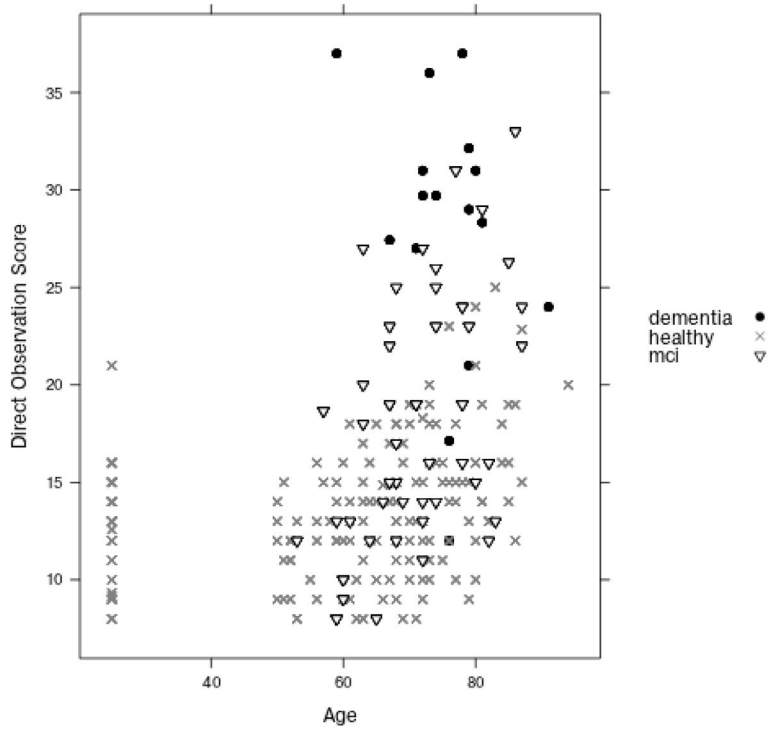


Figure 2. Distribution of the direct observation scores grouped by participant’s cognitive diagnosis. Participants are organized by age on the x axis and the y axis represents the corresponding score.

Date	Time	ID	Message
2010-01-27	09:04:07.001645	D007	OPEN Medicine-start
2010-01-27	09:04:07.004057	M016	OFF
2010-01-27	09:04:07.006439	M013	OFF
2010-01-27	09:04:09.009116	I006	ABSENT Medicine-step 2
2010-01-27	09:04:10.009602	M017	OFF
2010-01-27	09:04:11.058125	M017	ON
2010-01-27	09:04:13.004011	M017	OFF
2010-01-27	09:04:26.095202	I010	ABSENT Medicine-step 1
2010-01-27	09:04:29.006123	M017	ON Medicine-step 2
2010-01-27	09:04:31.050852	M017	OFF
2010-01-27	09:04:33.005186	P001	3439
2010-01-27	09:04:45.005352	P001	3428
2010-01-27	09:04:54.059162	M017	ON Medicine-step 2
2010-01-27	09:04:56.008738	M017	OFF
<hr/>			
2010-01-27	09:06:18.086175	I010	PRESENT
2010-01-27	09:06:20.011737	I010	ABSENT
2010-01-27	09:06:20.014137	I006	PRESENT Medicine-step 3
2010-01-27	09:06:24.046323	M017	ON Medicine-step 3
2010-01-27	09:06:25.001502	P001	1185
2010-01-27	09:06:26.000764	M017	OFF
2010-01-27	09:06:29.091058	I010	PRESENT Medicine-step 3
2010-01-27	09:06:33.000376	D007	CLOSE Medicine-end

Figure 3.

Sensor file format and sample annotation. Sensor IDs starting with M represent motion sensors, D represents door sensors, I represents item sensors, and P represents power usage sensors. The data is annotated with the start and end points of the activity (in this case, Medicine) and the individual step numbers within the activity.

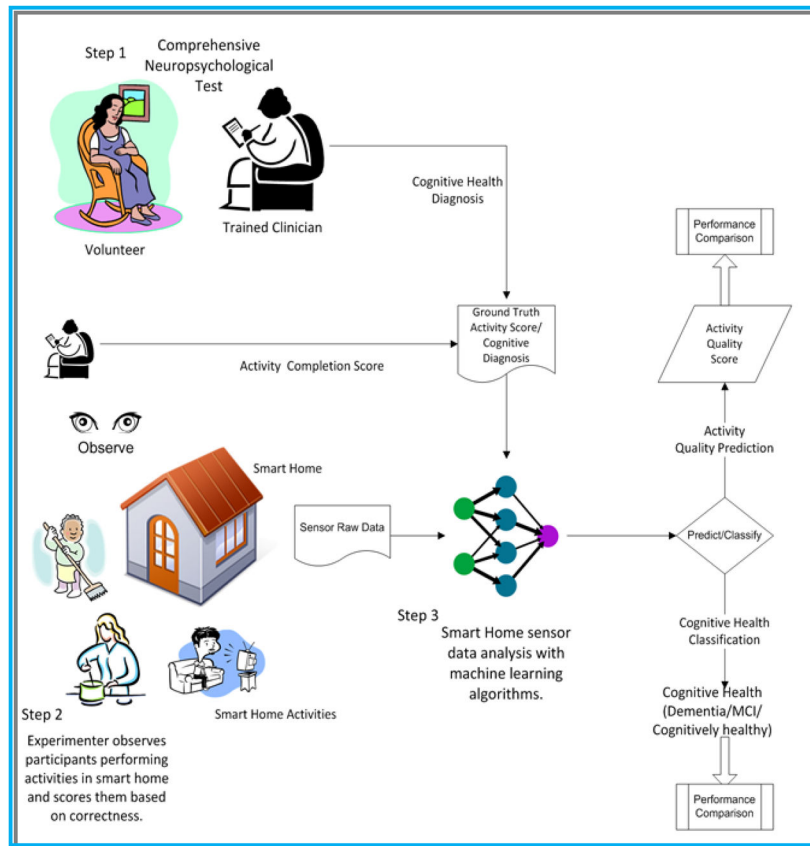


Figure 4. Steps involved in perform sensor-assisted cognitive health assessment. The process starts with a comprehensive neuropsychology assessment of the participant. The participant then performs IADLs in a smart home monitored by trained clinicians and smart home environmental sensors. The raw sensor data is annotated with activity labels. From the annotated sensor data, we extract features and analyze it with machine learning algorithms to derive the quality of the activity. The results are used by a clinician or by a computer program to perform cognitive health assessment.

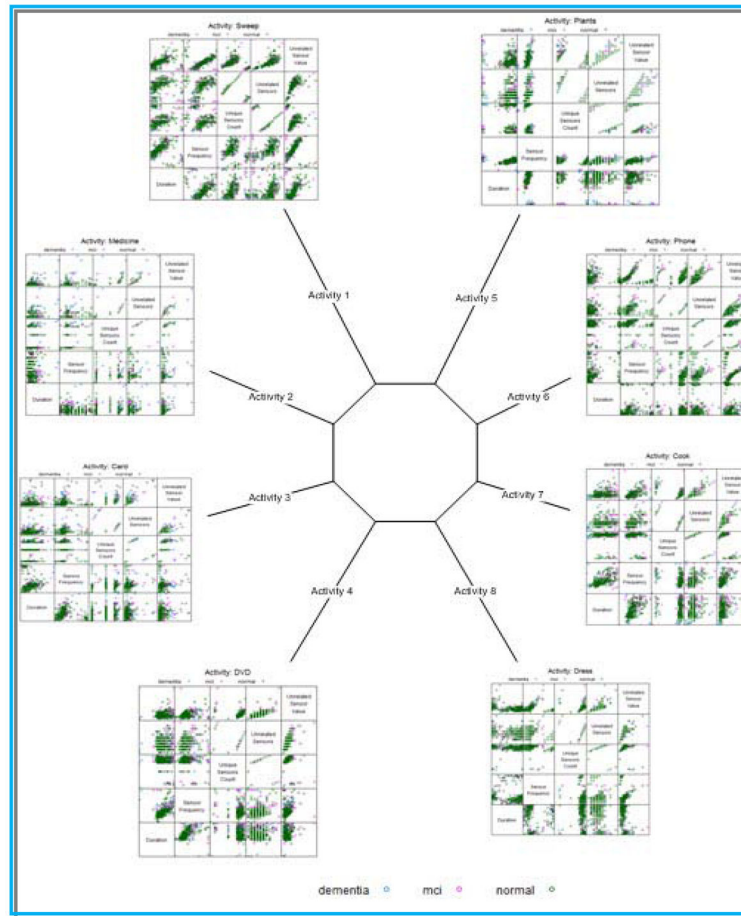


Figure 5.

Scatter plot of sensor features for each of the eight activities. Each grid cell in the plot represents a combination of two of the sensor features (duration, sensor frequency, unrelated sensors, and unrelated sensor count).

Table 1

Inclusionary and exclusionary criteria for the MCI, healthy older adult and dementia groups.

Initial screening for all middle age and older adult participants:
<ul style="list-style-type: none"> a. Medical interview to rule out exclusionary criteria of history of brain surgery, head trauma with permanent brain lesion, current or recent (past year) self-reported history of alcohol or drug abuse, stroke, or a known medical, neurological or psychiatric cause of cognitive dysfunction (e.g., epilepsy, schizophrenia). b. Clinical dementia rating instrument to assess dementia staging [35]. c. Telephone interview of cognitive status [36] to assess cognitive status and exclude significantly impaired participants who would be unable to complete the assessment.
Inclusion criteria for MCI group:
<ul style="list-style-type: none"> a. Self-report or knowledgeable informant report of subjective memory impairment for at least 6 months, as assessed by direct questioning during initial screening interview. b. Objective evidence of impairment in single or multiple cognitive domains (memory, executive, speeded processing, and/or language), with scores falling at least 1.5 standard deviations below age-matched (and education when available) norms. Test listed by domain with reference to norms used in parentheses: <ul style="list-style-type: none"> <i>Memory:</i> Memory Assessment Scale list learning and long- delayed free recall [42]; Brief Visual Memory Test learning and long-delayed free recall [41] <i>Executive:</i> Delis-Kaplan Executive Functioning Scale total correct from the Letter Fluency and Design Fluency subtests [43]; Trail Making Test, Part B total time [44]; Wechsler Adult Intelligence Scale-Third Edition Letter-Number Sequencing subtest total correct [46] <i>Speeded processing:</i> Trail Making Test, Part A total time [44]; Symbol Digit Modalities Test total correct written and oral [45] <i>Language:</i> Boston Naming Test total correct [44]; Delis-Kaplan Executive Functioning Scale Category Fluency subtest total correct [43] c. Preserved general cognitive functions as confirmed by a score of 27 or above on the TICS (equivalent to the normality cutoff score of 24 on the Mini Mental Status Exam) [48]. d. No significant impact of cognitive deficits on the participant's daily activities, as confirmed by a total CDR score of no greater than 0.5, which is consistent with minimal change in the participant's habits. e. Nonfulfillment of the <i>DSM-IV-TR</i> criteria for dementia (American Psychiatric Association, 2000), confirmed by reviewing screening data, neuropsychological testing data, and any available medical records.
Inclusion criteria for healthy older adult controls:
<ul style="list-style-type: none"> a. Reported no history of cognitive changes. b. Scored within normal limits on the TICS. c. Scored a 0 on the clinical dementia rating.
Inclusion criteria for dementia patients:
<ul style="list-style-type: none"> a. Met DSM IV-TR criteria for dementia [40] b. Scored 0.5 or higher on the clinical dementia rating.

Table 2

Coding scheme to assign direct observation scores to each activity.

Score	Criteria
1	Task completed without any errors
2	Task completed with no more than two of the following errors: non-critical omissions, non-critical substitutions, irrelevant actions, inefficient actions
3	Task completed with more than two of the following errors: non-critical omissions, non-critical substitutions, irrelevant actions, inefficient actions
4	Task incomplete, more than 50% of the task completed, contains critical omission or substitution error
5	Task incomplete, less than 50% of the task completed, contains critical omission or substitution error

Table 3

Sensor-based feature descriptors for a single activity.

Feature #	Feature Name	Feature Description
1	Age	Age of the participant
2	Help	An indicator that experimenter help was given so that the participant could complete the task
3	Duration	Time taken (in seconds) to complete the activity
4	Sequence length	Total number of sensor events comprising the activity
5	Sensor count	The number of unique sensors (out of 36) that were used for this activity
6..31	Motion sensor count	A vector representing the number of times each motion sensor was triggered (there are 26 motion sensors)
32	Door sensor count	Number of door sensor events
33	Item sensor count	Number of item sensor events
34	Unrelated sensors	Number of unrelated sensors that were triggered
35	Unrelated sensor count	Number of unrelated sensor events
36	Health status	Status of the patient: Healthy, MCI, or Dementia

Table 4

Pearson correlation between activity sensor-based scores and activity direct observation scores for sample subsets using SVM. For each sample subset, there are eight different learning models, each of which learns a mapping between a single activity and the corresponding direct observation score. The samples are cognitive healthy (CH) participants, participants with mild cognitive impairment (MCI) and participants with dementia (D) (* $p < 0.05$, ** $p < 0.005$, † $p < 0.05$ with Bonferroni correction for the three sample groups).

Sample	N	Pearson Correlation coefficient (r)							
		Sweep	Medicine	Card	DVD	Plants	Phone	Cook	Dress
CH	196	0.50**†	0.02	0.04	0.22**†	-0.04	0.31**†	0.18**†	0.22**†
MCI	51	0.58**†	-0.01	0.07	0.13	0.01	0.18	0.35**†	0.03
CH,MCI	247	0.58**†	0.08	-0.12	0.24**†	0.05	0.33**†	0.31**†	0.24**†
CH,D	212	0.58**†	0.16**†	-0.09	0.24**†	0.08	0.31**†	0.28**†	0.22**†
MCI,D	67	0.75**†	0.01	0.21	0.03	0.38*	0.02	0.32**†	0.05
CH,MCI,D	263	0.63**†	0.17**†	-0.07	0.27**†	0.09	0.33**†	0.37**†	0.23**†

Table 5

Pearson correlation and Spearman rank correlation between the summed sensor-based scores and direct observation scores for sample subsets using SVM. Samples are cognitive healthy (CH) participants, participants with mild cognitive impairment (MCI), and participants with dementia (D) (* $p < 0.05$, ** $p < 0.005$, † $p < 0.05$ with Bonferroni correction for the three sample groups). The first correlation coefficient listed is the Pearson correlation coefficient while the second value is the Spearman rank correlation coefficient.

Sample	N	$r(\text{all features})$	$r(\text{sensor features})$	$r(\text{sensor features without duration})$
CH	196	0.39**†	0.22**†	0.20**
		0.42**†	0.25**†	0.22**
MCI	51	0.50**†	0.35*	0.26*
		0.48**†	0.31*	0.20
D,CH	212	0.50**†	0.47**†	0.46**†
		0.48**†	0.39**†	0.39**†
MCI,CH	247	0.49**†	0.34**†	0.32**†
		0.48**†	0.32**†	0.30**†
MCI,D	67	0.59**†	0.60**†	0.53**†
		0.63**†	0.60**†	0.52**†
CH,MCI,D	263	0.54**†	0.51**†	0.49**†
		0.52**†	0.44**†	0.43**†

The column $r(\text{all features})$ lists correlation coefficients obtained using all features, $r(\text{sensor features})$ lists correlation coefficients obtained using only sensor based features, and $r(\text{sensor features without duration})$ lists correlation coefficients obtained using all sensor based features without the duration feature.

Correlation between activity sensor-based scores and activity direct observation scores for sample subsets using principal component analysis (PCA). We standardized the data before applying PCA. Samples are cognitive healthy (CH) participants, participants with mild cognitive impairment (MCI), and participants with Dementia (D) (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, † $p < 0.05$ with Bonferroni correction for the three sample groups).

Table 6

Sample	Size(N)	Pearson Correlation coefficient (r)									
		Sweep	Medicine	Card	DVD	Plants	Phone	Cook	Dress		
CH	196	-0.30***†	0.09	0.06	0.23***†	0.04	0.03	0.26***†	-0.16*		
MCI	51	-0.51***†	0.14	0.08	-0.17	0.38**†	0.20	0.11	-0.15		
CH,D	212	-0.37***†	0.26***†	-0.02	0.22***†	0.06	0.04	0.18***†	-0.14*		
CH,MCI	247	-0.36***†	0.13	0.00	0.25***†	0.20***†	0.10	0.23***†	0.08		
MCI,D	67	-0.58***†	0.18	0.15	-0.11	0.20	0.19	-0.03	0.12		
CH,MCI,D	263	-0.41***†	0.23***†	-0.05	0.23***†	0.15**†	0.10	0.17**†	0.07		

Table 7

Pearson correlation and Spearman rank correlations between the summed sensor-based scores and direct observation scores for sample subsets using principal component analysis. We standardize data before applying PCA. The samples are cognitive healthy (CH) participants, participants with mild cognitive impairment (MCI), and participants with dementia (D) (* $p < 0.05$, ** $p < 0.005$, † $p < 0.05$ with Bonferroni correction for the three sample groups). The first value listed is the Pearson correlation coefficient while the second value is the Spearman rank correlation coefficient.

Sample	Sample size	$r(\text{all features})$	$r(\text{sensor features})$	$r(\text{sensor features without duration})$
CH	196	0.13	0.12	0.11
		0.12	0.10	0.08
MCI	51	0.10	0.08	0.09
		-0.03	-0.08	-0.08
D,CH	212	0.13	0.11	0.10
		0.12	0.09	0.08
MCI,CH	247	0.18**†	0.16**†	0.15*
		0.08	0.07	0.05
MCI,D	67	0.12	0.10	0.10
		0.09	0.08	0.07
CH,MCI,D	263	0.16**†	0.14*	0.13
		0.07	0.05	0.03

The column $r(\text{all features})$ lists correlation coefficients obtained using all features, $r(\text{sensor features})$ lists correlation coefficients obtained using only sensor based features, and $r(\text{sensor features without duration})$ lists correlation coefficients obtained using all sensor based features without the duration feature.

Table 8

AUC (first entry) and G-mean (second entry) values for automated support vector machine of cognitive health status for each activity.

Sample	Costs	Sweep	Medicine	Card	DVD	Plants	Phone	Cook	Dress
D, MCI	(3, 1)	0.70*	0.64	0.63	0.63*	0.68	0.54	0.78*	0.66*
		0.69	0.61	0.51	0.61	0.55	0.52	0.75	0.68
MCI, CH	(5, 1)	0.60	0.68	0.60	0.58	0.57	0.63	0.64	0.65
		0.57	0.64	0.62	0.57	0.57	0.62	0.56	0.59
CH, D	(23, 1)	0.82*	0.67*	0.81*	0.60*	0.76*	0.63	0.87*	0.76*
		0.79	0.69	0.81	0.60	0.79	0.57	0.73	0.69

* Classifier with better than weighted random prediction

AUC (first entry) and G-mean (second entry) values for automated support logistic regression classification of cognitive health status for each activity.

Table 9

Sample	Costs	Sweep	Medicine	Card	DVD	Plants	Phone	Cook	Dress
D, MCI	(3, 1)	0.71*	0.53	0.44	0.64	0.59	0.53	0.86*	0.63
		0.68	0.58	0.45	0.62	0.58	0.54	0.75	0.61
MCI, CH	(5, 1)	0.61*	0.71*	0.60*	0.58	0.60*	0.64*	0.60	0.65*
		0.59	0.64	0.56	0.52	0.55	0.58	0.55	0.60
CH, D	(23, 1)	0.77	0.57	0.73	0.61	0.74	0.63	0.93*	0.66*
		0.66	0.48	0.58	0.63	0.50	0.57	0.80	0.67

* Classifier with better than weighted random prediction

Table 10

Combined cost-sensitive health classification performance with all activities classified using a SVM classifier.

Sample	Sample size	Costs	AUC	G mean
MCI,D	67	(2, 1)	0.56	0.43
D,CH	212	(5, 1)	0.72	0.65

Table 11

Combined cost-sensitive health classification performance with all activities classified using a logistic regression classifier.

Sample	Sample size	Costs	AUC	G mean
MCLD	67	(2, 1)	0.53	0.40
MCLCH	247	(5,1)	0.66	0.62
D,CH	212	(5, 1)	0.83	0.75

Table 12

Combined cost-sensitive health classification with selected activities classified using a SVM classifier.

Sample	Sample size	Costs	AUC	G mean
MCI,D	67	(2, 1)	0.59	0.53
D,CH	212	(6, 1)	0.80	0.73

Table 13

Combined cost-sensitive health classification with selected activities classified using a logistic regression classifier.

Sample	Sample size	Costs	AUC	G mean
MCLD	67	(3, 1)	0.62	0.54
MCLCH	247	(3,1)	0.65	0.60
D,CH	212	(3, 1)	0.87	0.75