Behavioral Differences Between Subject Groups Identified Using Smart Homes and Change Point Detection

Gina Sprint, *Member, IEEE*, Diane J. Cook, *Fellow, IEEE*, and Roschelle Fritz

***Abstract*—With the arrival of the internet of things, smart environments are becoming increasingly ubiquitous in our everyday lives. Sensor data collected from smart home environments can provide unobtrusive, longitudinal time series data that is representative of the smart home resident's everyday behavior and its changes over time. When long term behavioral data is available from multiple smart home residents, differences between groups of subjects can be investigated. Group-level discrepancies can offer insight into isolating behaviors that manifest in our everyday routines due to a health concern or major lifestyle change. To acquire such insights, we propose an algorithmic framework based on change point detection called Behavior Change Detection for Groups (BCD-G). With BCD-G we hypothesize that we can quantify and characterize differences in behavior between groups of individual smart home residents. We evaluate our BCD-G framework with one month of daily sensor data collected from two groups of seven smart home residents each. All subjects in the first group exhibit a diagnosed cognitive impairment. The remaining seven residents form a second group of cognitively healthy, age-matched controls. Using BCD-G, we identify differences between these two groups, such as how their health impairment affects their patterns of performing activities of daily living and how clinically relevant behavioral features, such as in-home walking speed, differ due to the impairment. With the unobtrusive monitoring of smart home environments, clinicians can use BCD-G for remote identification of behavior changes indicative of early onset of health concerns.**

***Index Terms*—Activity recognition, smart environments, change point detection, unsupervised machine learning, remote health monitoring**

I. Introduction

Sensor technology is becoming a commonplace aspect of our everyday lives. With the internet of things, we wear sensors on our bodies, and we install sensors in our environments. Specifically for smart environments, sensors are often installed in our residence to create a smart home, in our workplace to create smart offices, and even on and around our streets to create smart cities. These sensors continuously and unobtrusively collect time series data about our everyday behavior, both in and out of the home. Such time series data consist of sensor state information along with a timestamp for when the state change occurred. Using activity recognition and machine learning, time series data can be analyzed to detect changes in our everyday behavior over short or long periods of time [1], [2]. Often changes in the daily behavior patterns of an individual are attributed to both major and minor changes in health [3], [4], such as breaking a bone or catching a cold. Health changes can abruptly begin with an isolated event, such as a fall in the home, or can slowly manifest over time [5], such as the onset of dementia. For all such health changes, sensor data collected from smart environments can provide unprecedented, naturalistic insight about how different groups and individuals experience the onset of and recovery from a health concern.

In this paper, we propose an algorithmic framework for comparing the daily behavior patterns of groups of individuals living in smart environments. Our framework is called Behavior Change Detection for Groups (BCD-G), and it is an expansion of our prior work, BCD, that analyzes changes within time series collected from a single smart home resident [6], [7]. BCD and BCD-G are change point detection-based algorithmic frameworks for detecting and analyzing changes in time series data. Change point detection algorithms compute a change score value that quantifies how much change is surrounding a candidate point in time. With BCD-G, we propose a novel approach to detect and compare the differences in behavior between groups of individuals by treating a health or behavioral trait like a candidate change point. We hypothesize that these differences can provide behavioral insights into a specific health trait that will allow earlier detection and recovery. To apply BCD-G and compare groups, we need at least two groups. The first group contains individuals with the same health/behavioral trait that is under investigation. The second group is a similarly sized group of age-matched individuals who do not exhibit the same health/behavioral trait as the first group. This second group serves as a control group. We then apply BCD-G to these two groups to provide a description of the

*Paper submitted for review on January 10, 2020. This work was supported in part by the National Institutes of Health under Grants R01EB009675, R01NR016732, 25EB024327, and R25AG046114 and by the National Science Foundation under Grant 1734558.

G. Sprint is with Gonzaga University, Spokane, WA 99258 USA. (e-mail: sprint@gonzaga.edu).

D. J. Cook is with Washington State University, Pullman, WA 99164 USA. (e-mail: djcook@wsu.edu).

R. Fritz is with Washington State University Vancouver, Vancouver, WA 98686 USA. (e-mail: shelly.fritz@wsu.edu).

everyday behavior that constitutes the health/behavioral trait under scrutiny.

We evaluate our BCD-G approach with two groups of seven individuals, each living in smart homes. All smart home residents in the first group have a cognitive impairment and are diagnosed with mild cognitive impairment (MCI) or mild dementia. The second group consists of cognitively healthy smart home residents who are age-matched to the first group. Using these two groups as input to our BCD-G framework, we are able to describe behavioral differences that manifest in cognitively impaired individuals. More specifically, our results describe differences in terms of common activities of daily living, mobility, and sleeping patterns. This information can be leveraged by clinicians, caretakers, and by the residents themselves to detect an early onset of a medical condition and enable even earlier preventative care.

II. Related Work

Researchers have long investigated the connection between behavior and health. While in the past, such investigations have relied on observation and self-report, the recent ability to unobtrusively collect large amounts of sensor data opens the pathway to quantifying and understanding this connection in an ecologically valid manner. Sensors provide information on a wide array of physiological and behavioral features. In recent years, these sensors have become low cost, wireless, seamlessly integrated into multi-purpose systems, and deployable in real-world settings.

One popular sensing modality is wearable sensors. Wearable sensors, embedded in mobile phones, smartwatches, and even clothes, are traditionally used to monitor and track movement based activities such as sit, stand, walk, run, and lie down. Using these mobile packages, researchers have been able to quantify the relationship between cognitive health and such behavior factors as sleep [8], [9], gait [10], [11], time out of the home [12], and amount of phone-based social interaction [13]. These studies represent an important foray into the area of sensor-based behavior analysis, yet they focus on a single aspect of behavior.

In other studies, researchers assessed multiple behavior parameters in home settings using unobtrusive ambient sensors. While most of these studies were performed in controlled conditions with scripted activities, significant correlations were discovered between behavioral factors and traditional neuropsychological test scores [14]–[16]. These behavior parameters also correlated with other health components, including fall risk, cognitive function, motor function, and dyskinesia "on" states [15]–[19]. A few more recent studies monitored individuals in their home environments over multiple months or years [20], [21]. In these cases, walking speed, time out of the home, time spent in specific home locations, and variation in daily routine were predictors of cognitive health [21]–[24].

One aspect that characterizes most of these studies is that scientists sought to understand the connection between a specific behavioral characteristic and a component of cognitive or physical health. In contrast, the study we present in this paper takes the opposite view. We collect sensor-based behavior data for individuals from multiple subgroups. We then introduce data mining techniques that find the behavior differences that can be used to predictively characterize the health differences between the groups. This way, we can quantify the amount of differences in overall routine that exists between individuals within the same diagnosis category and between different diagnosis categories. Furthermore, we can characterize the nature of these differences. These insights can be used to better understand the behavior impacts of cognitive impairment and to inform the design of assessment measures.

TABLE I
SMART HOME RESIDENT PARTICIPANT CHARACTERISTICS.

Group	Participant ID	Age (years)	Education (years)	Health problems
Cognitive Impairment (CI)	hh101	87	16	MCI
	hh104	83	16	Dementia
	hh116	79	20	MCI
	hh118	83	16	MCI
	hh119	81	18	MCI
	hh122	82	18	MCI, early dementia
	hh123	89	18	Dementia
Healthy Control (HC)	hh103	79	20	N/A
	hh105	83	16	N/A
	hh106	73	16	N/A
	hh108	80	16	N/A
	hh109	90	16	N/A
	hh111	81	20	N/A
	hh114	93	12	N/A

MCI = mild cognitive impairment.

III. Methods

In this paper, we analyze smart home data collected from 14 single-resident apartments instrumented with the CASAS "smart home in a box" [25]. Each CASAS smart home apartment has ceiling-installed motion/light sensors and door/temperature sensors. The door/temperature sensors are attached to the apartment entry doors and to commonly used cabinets (such as kitchen and bathroom cabinets). The CASAS system logs events from each sensor when an update occurs, such as a state change from "motion" to "no motion" for a motion sensor or "open" to "closed" for a door sensor. For each of the 14 smart home residents, we analyzed one month (30 days) of continuously and unobtrusively collected sensor event data. We selected these 14 smart home residents for analysis in order to form two groups of 7 residents each, the "cognitive impairment" group, denoted CI, and the "healthy" control group, denoted HC. The cognitive impairment group consists of residents who were diagnosed with MCI or dementia before data collection began. The groups are age-matched and participants in both groups exhibit similar education levels. A summary of the participants in each group is listed in Table I.

A. Activity Recognition

To provide context for what the resident is doing in the smart home, we use activity recognition (AR) to label each sensor event with an activity label. To do this, we use the CASAS-AR algorithm which has demonstrated high labeling accuracy in

our previous work [26]. The CASAS-AR algorithm assigns each sensor event with one of 40 activity labels, such as "Eat", "Enter Home", and "Sleep". Ground truth labels are provided by external staff. Annotators label each sensor data reading with a corresponding activity label based on the home's floor plan, sensor layout, and a resident-provided description of the common times and locations for routine activities.

Many of the activity labels represent pre-defined activities of daily living (ADLs), while other labels are based on individual smart home resident behavior. There is one activity label, "Other Activity", that is assigned when the detected resident behavior does not match an ADL or commonly known behavior. Because not all participants performed all 40 activities at some point during the month of data collection, in a post-processing step we combine specific activity labels into more general activity labels to form a common activity set. To do this, we combine activities based on similarity. For example, "Cook", "Cook Breakfast", "Cook Dinner", and "Cook Lunch" are all combined into the single activity label "Cook." This reduces the initial set of activities from 40 to 18 labels. We then remove any labels from this set of 18 general labels if a participant did not perform the activity at least once during the one month of data we are analyzing. There are two activities, "Exercise" and "Housekeeping", that met the criteria and are removed from analysis. Our final activity label set contains the following 16 labels (with the combined activity labels for a final label in parentheses):

1. Bathe
2. Bed to Toilet Transition
3. Cook (Cook, Cook Breakfast, Cook Dinner Cook Lunch)
4. Dress
5. Eat (Eat, Eat Breakfast, Eat Dinner, Eat Lunch Drink)
6. Enter Home
7. Leave Home (Leave Home, Step Out)
8. Meds (Take Medicine, Evening Meds, Morning Meds)
9. Other Activity
10. Personal Hygiene (Personal Hygiene, Groom)
11. Relax (Relax, Watch TV, Read)
12. Sleep (Sleep, Sleep Out of Bed, Go to Sleep, Nap, Wake Up)
13. Socialize (Entertain Guests, Phone)
14. Toilet
15. Wash Dishes (Wash Dishes, Wash Breakfast Dishes, Wash Dinner Dishes, Wash Lunch Dishes)
16. Work (Work, Work at Desk, Work at Table, Work on Computer)

B. Feature Extraction

Using the activity-labeled time series data, we perform several transformations of the data to prepare it for input to our BCD-G algorithm. First, we compute the probability of each activity occurring at each hour of the day. These hourly probabilities formed a multivariate time series of size 24D x 16, where D is the number of days of analyzed smart home data. For our analysis, D=30 days because we are analyzing one month of daily behavior data. Secondly, we extract features from the smart home event and activity data in order to provide context for analyzing behavioral changes. We utilize features that quantify daily behavior and are straightforward to interpret.

TABLE II
FEATURES EXTRACTED FOR CHANGE ANALYSIS

Type	Daily Feature
Activity	Total duration of each activity
	First occurrence of each activity (measured as seconds past midnight)
	Last occurrence of each activity (measured as seconds past midnight)
Sleep	Total duration of sleep during the day
	Total duration of sleep during the night
	Number of nighttime sleep interruptions (measured as number of non-sleep activity sequences between sleep sequences during the night)
Mobility	Total movement in the home (measured as distance in feet)
	Average walking speed (measured as feet per second during bouts of movement)
Routine	Circadian rhythm strength (ratio of nighttime duration of activity (e.g. non-sleep and non-relax activities) divided by the previous day daytime duration of activity) [27]
	Complexity of routine (measured as entropy of activity probabilities)
	Number of different daily activities
	Variability in activity durations (measured as the standard deviation of activity sequence durations)

Within a 24-hour period, we define the day subperiod to be 7:00am to 6:59pm. We define the night subperiod to be 7pm to 6:59am. Specifically for nighttime sleep, we calculate the sum of sleep during two periods: the first period beginning the previous day at 7:00pm and ending at 11:59pm and the second period beginning on the current day at 12:00am and ending at 5:59am. In total, we extracted 57 features describing daily behavior. These features are summarized in Table II.

C. Behavior Change Detection

With the aforementioned hourly activity probabilities and daily behavior features, we aim to describe the differences between the CI and the HC groups using change point detection. In our prior work, we proposed a window-based change detection approach called behavior change detection (BCD) [6], [7]. BCD accepts time series data from a single participant as input, such as a year of AR-labeled smart home data or a month of wearable sensor data. Based on configured algorithm parameters for processing the time series, BCD outputs detected changes between periods of the time series data, called windows. More specifically, the output consists of a list of change score values computed between equal-size windows of the data (e.g. daily windows, weekly windows, monthly windows). For large change scores representing a significant change between two windows, BCD also provides information to interpret the source of the detected change. To summarize, BCD is comprised of four main steps:

1. Segment the data into windows
2. Compute change scores between windows
3. Test the significance of the detected changes
4. Analyze the source of significant changes

BCD is a framework that supports alternative change detection algorithms to be "plugged in" for step #2. In our previous work, we utilized several different algorithms, including Relative Unconstrained Least-Squares Importance Fitting (RuLSIF) [28], Permutation-based Change Detection in Activity Routine (PCAR) [29], small window adaption of PCAR (sw-PCAR) [30], texture dissimilarity [31], [32], and virtual classifier (VC) [33]. In this paper, we utilize the PCAR

and VC algorithms, as well as a new change detection algorithm for BCD, SEParation change point detection (SEP) [34]. We use the PCAR algorithm because it was designed for analyzing longitudinal smart home data. We use the VC algorithm because it provides a test for significance based on the binomial distribution that is applied to step #3 of BCD, as well a decision tree that is informative for step #4 of BCD. We include SEP because it is a new change point detection algorithm that has demonstrated high performance on similar smart home time series data [34], [35]. SEP also supports multi-dimensional time series data, which allows us to apply SEP to daily feature matrices. In the following three sections, we briefly describe each of these three change point detection algorithms.

1) Permutation-based Change Detection in Activity Routine

The PCAR algorithm detects changes in smart home data using an activity curve model. The activity curve model represents activity probability distributions for each time interval in a day (m time intervals per day), aggregated over a window of n days. The algorithm detects change between two activity curves by first aligning the two curves using dynamic time warping (DTW). A distance function is used to quantify the dissimilarity between each pair of time intervals between the two DTW-aligned curves. To produce a single change score from this vector of m distances, the two windows are concatenated to form a window of $2n$ total days. The activity curve extraction, the DTW-alignment, and the distance computation process are repeated N times for random permutations of the days in the concatenated windows. This procedure produces a $N \times m$ permutation matrix. The ratio of permutation-based distances exceeding the original distance is then computed for each time interval, producing a list of ratios. These ratios serve as a p-values representing the probability the original distance was sampled from the same distribution produced by the permutations for a given time interval. A Benjamini-Hochberg correction is then performed for a given α ($\alpha < 0.05$) to account for the multiple comparisons performed [36]. The last step of the PCAR algorithms is to count the remaining p-values that are significant at the given α level. This count is the PCAR change score representing the change between the two activity curves.

2) SEParation Change Point Detection

A group of non-parametric change point detection algorithms, called likelihood ratio estimators, compute a change score by estimating the ratio of probability distributions between two time series windows surrounding a suspected change point. The higher the ratio, the more likely the windows are different and there is a valid change point between them. One of the most common likelihood ratio estimation algorithms is called RuLSIF [28]. RuLSIF uses the Pearson divergence dissimilarity measure to estimate the probability distribution. Like RuLSIF, we introduced SEP in prior work as a non-parametric change point detection algorithm [34]. Instead of the Pearson divergence dissimilarity measure, SEP uses a separation distance function. Both RuLSIF and SEP require computing a parameter threshold value to determine if a computed change score is significant or not based on whether it is above or below the threshold. We are interested in using SEP over RuLSIF because of its increased performance in prior work and because it supports both one-dimensional and two-dimensional time series data [34], [35]. Furthermore, we found in prior work that SEP outperformed RuLSIF in detecting true behavior transitions from smart home data [35].

3) Virtual Classifier

Hido and colleagues proposed the VC algorithm as part of their change analysis framework [33]. Change analysis computes a change score, while providing context to identify the features that contributed to the change. The approach begins with two $n \times z$ feature matrices, where n is the time series window size (e.g. a number of days) and z is the number of features extracted from each day in the window. Feature vectors (rows) in the first matrix are labeled with a hypothetical "positive" class, while feature vectors in the second matrix are labeled with a hypothetical "negative" class. A decision tree is trained using k-fold cross validation to distinguish between these two classes. The average classification accuracy is then compared to a significance threshold, which is based on the expected accuracy of a binary classifier trained on random samples. Using the binomial distribution's inverse survival function, a probability value p_{critical} is computed using the two window sizes and a significance level, α ($\alpha < 0.05$). If the VC accuracy is greater than or equal to p_{critical}, then the VC accuracy is considered to be significant [33]. For example, if the two window sizes are both five days and $\alpha = 0.05$, then using the inverse survival function of the binomial distribution, $p_{\text{critical}} = 0.8$. This would mean a VC accuracy acquired by training using only 10 days would have to be greater than or equal to 0.8 in order to be determined significant. If the VC accuracy is significant, then retraining a decision tree on the entire dataset produces a tree with the features most responsible for the change between the two windows towards the top of the tree.

D. BCD for Group Analysis

In previous work, we developed BCD to find points in time with detected behavior changes that could be indicative of health events [6]. In the current paper, we expand our BCD framework to compare populations (e.g. groups) instead of comparing behavior for one individual at different time points. To describe BCD-G, let G_1 and G_2 represent behavior features sampled from two different groups. Our goal is to compare behavior between the groups and characterize the differences. Let G_1 contain behavior data for N individuals while G_2 contains behavior data for M participants. The changes between G_1 and G_2, denoted $G_1 \Delta G_2$, can be quantified using BCD-G by setting the "windows" to be "participants". This is in contrast to standard BCD where windows are sub-samples of time series collected from the same participant. A comparison of two windows in BCD-G is a comparison of two participants. We set up BCD-G to perform comparisons in two different configurations: group-to-group and pair-to-pair.

1) Group-to-Group Comparisons

Suppose G_1 and G_2 both contain the same number of participants. We concatenate equal-length time series for all participants, $p \in G_1$, to form a new time series window, W_1. This process is repeated for all participants $q \in G2$ to form W_2. W_1 and W_2 are then concatenated to form a time series that is input to BCD-G with window size set to $|W_1|$, where $|W_1| = |W_2|$. Each day in the W_1 window is labeled as a positive class, while each day in the W_2 window is labeled with a negative class. Figure 1 shows an example of these two groups, each containing seven participants. A single comparison is made

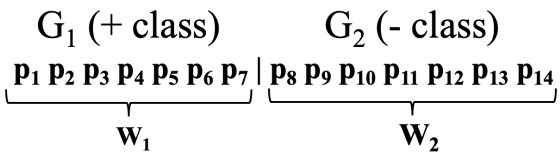
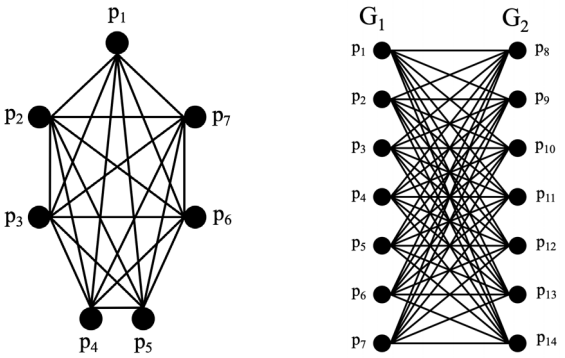


Fig. 1. Window representation of group-to-group comparisons between two groups, G_1 and G_2, each containing 7 participants. Each p_i is an equal-length behavior time series collected from participant i.



a) Intragroup pairs of a group with 7 participants. The graph is complete.

b) Intergroup pairs of two groups G_1 and G_2, each with 7 participants. The graph is complete bipartite.

Fig. 2. Graphical representation of pair-to-pair comparisons.

between W_1 and W_2 using a virtual classifier trained to discriminate between positive and negative days.

2) Pair-to-Pair Comparisons

We compute change scores for all possible pairs of participants within each group and between groups. To elaborate, there are three distributions of pairwise change scores for two groups G_1 and G_2, intragroup pairwise change scores and intergroup pairwise changes scores. Intragroup pairwise change scores are computed between all possible pairs of participants within a group G (see Figure 2a for a diagram of pairs in a general group). If there are multiple groups, as is the case in this paper with the two groups CI and HC, intragroup pairwise change scores are computed separately for each group. Intergroup pairwise change scores are computed between all possible pairs of participants in G_1 and G_2; one participant in the pair from the G_1 group and the other participant in the pair from the G_2 group (see Figure 2b for a diagram of pairs).

IV. Results

In this paper, we are interested in the differences between our two groups of participants, which are labeled as follows:

1. G_1: CI (Cognitive Impairment group): 7 participants with mild cognitive impairment or mild dementia
2. G_2: HC (Healthy Control group): 7 healthy, age-matched participants

Additionally, we denote the intergroup change scores between CI and HC with the label IN. We use BCD-G to compute the changes CIΔHC in two different configurations: group-to-group VC change scores and pair-to-pair participant change scores. The results can be further categorized into comparisons using all activities and comparisons using individual activities.

A. Group-to-Group VC Comparisons

First, we execute BCD-G looking at the changes between cognitively impaired and control participants using all of the activity time series data and using all of the extracted features. We perform a group-to-group comparison of CI to HC using BCD-G with VC. All 30-day time series in the CI group are concatenated to form window W_1 and all 30-day time series in the HC group are concatenated to from window W_2. The window size was set to 210 days (30 days multiplied by 7 participants in each group). The group-to-group CIΔHC VC change score is 0.826, which is a significant result because it is greater than the significance threshold of $p_{\text{critical}} = 0.541$ (calculated with $\alpha = 0.05$). The resulting VC decision tree is shown in Figure 3.

While change scores using all activities provides a gross overview of changes between groups, zooming in to investigate the changes in each individual activity provides a closer look at the source of the changes. For each of the 16 activities, we apply BCD-G to investigate which activities had the greatest differences between the cognitively impaired and healthy control groups. We utilize BCD-G with the VC approach to compare participants in each group, but for this comparison we use only features related to each activity. For example, for the Cook activity, the features included are Cook duration, time of Cook first occurrence, and time of Cook last occurrence. Table III lists each activity's VC change score and whether the score is significant or not. Only the VC change score for the Other Activity label does not exceed the significance threshold ($p_{\text{critical}} = 0.541$).

B. Pair-to-Pair Comparisons

Next we quantify the CIΔHC change within groups and the change between groups at an individual subject level. To do this, we compute change scores for all possible pairs of participants within the CI and HC groups and between these two groups. To elaborate, there are three distributions of pairwise change scores (91 total pairs):

- Intragroup pair-to-pair change scores (CI and HC, 42 pairs): computed as all possible combinations of two participants sampled from a group of seven participants ($2 * {}_7C_2 = 42$ pairs).
- Intergroup pairwise change scores (IN, 49 pairs): computed between pairs of 14 participants, one participant in the pair from the CI group and the other participant in the pair from the HC group ($7 * 7 = 49$ pairs).

Initially, we input all hourly probability time series data for all activities to PCAR. For SEP and VC, we included all features. The BCD-G results for all features are shown in Figure 4. Figures 4a-c show the distribution of change scores for each group (within CI, within HC, and within both groups) shown as box plots. Figures 4d-f show the individual mean pairwise change scores for each participant in the CI group as bars with standard deviations shown as error lines. For reference, percentile-based control group scores are shown as horizontal lines.

Lastly, we perform pairwise comparisons again, but this time utilizing only data related to each individual activity. Of the 48 algorithm configurations (16 activities for each of the three algorithms, respectively), we include results for the three activities with the most distinct observed changes for each algorithm. These are the Bed to Toilet Transition PCAR change

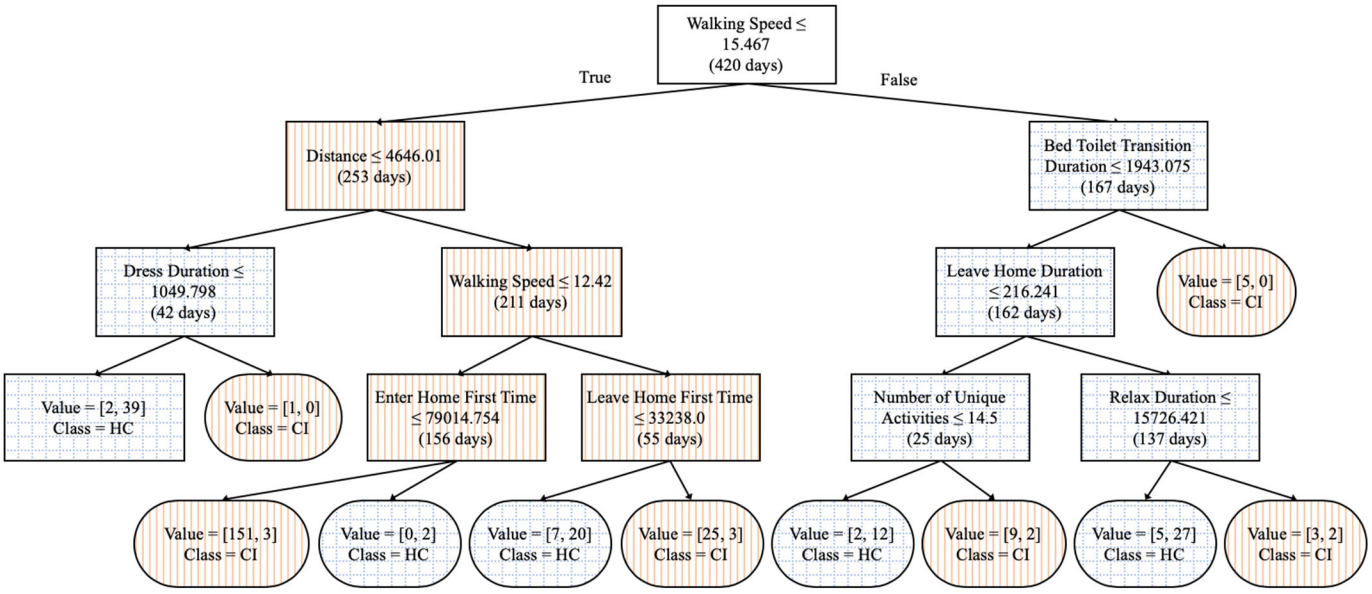


Fig. 3. Top VC decision tree rules for group-to-group comparisons using all activities. Rectangles are decision nodes and rounded rectangles are leaf nodes. Left branches are true and right branches are false. Orange vertical hatch represents participant subgroups containing mostly individuals from the cognitive impairment (CI) group and blue cross hatch represents participant subgroups containing mostly individuals from the healthy control group (HC).

TABLE III.
GROUP-TO-GROUP CIΔHC VC CHANGE SCORES

Activity	Virtual Classifier Score
Bathe	0.595*
Bed to Toilet Transition	0.600*
Cook	0.581*
Dress	0.710*
Eat	0.610*
Enter Home	0.695*
Leave Home	0.698*
Meds	0.695*
Other Activity	0.529
Personal Hygiene	0.652*
Relax	0.645*
Sleep	0.633*
Socialize	0.571*
Toilet	0.650*
Wash Dishes	0.643*
Work	0.581*

* denotes a significant change score with significance threshold $p_{\text{critical}} = 0.541$.

scores, Bathe SEP change scores, and Enter Home VC change scores. These results are shown as box plots and bar charts in Figure 5.

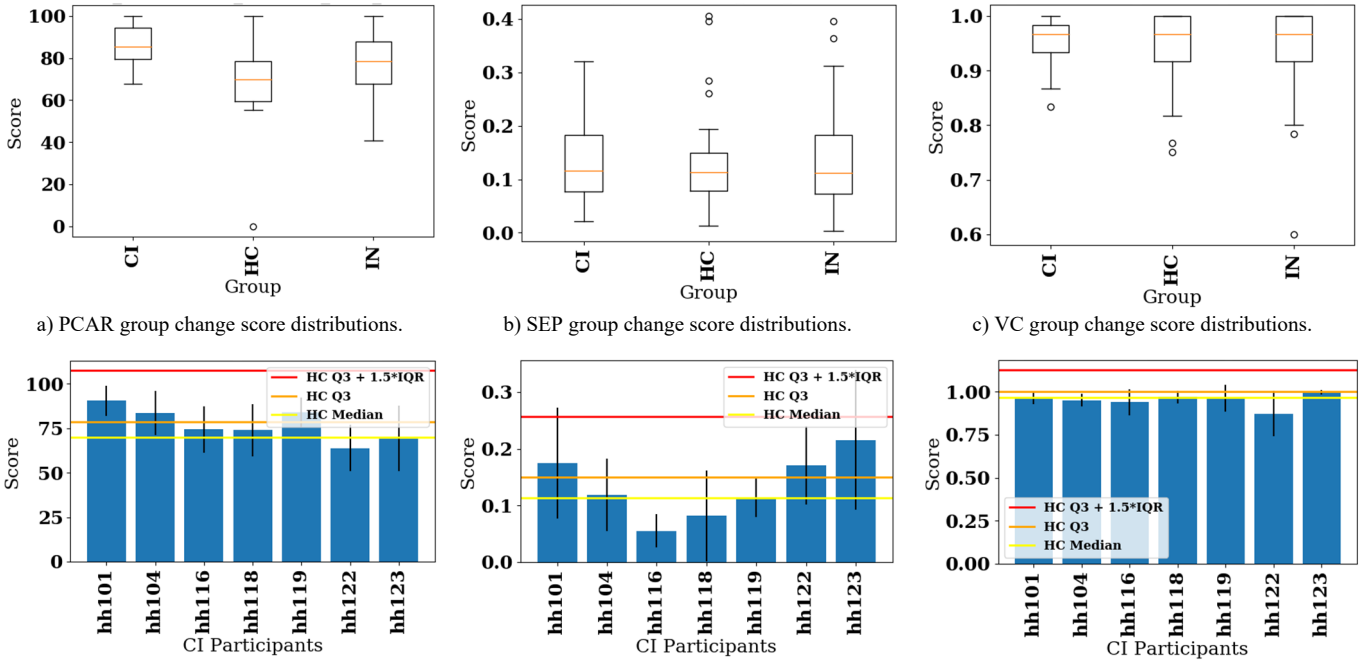
V. Discussion

In this paper, we adapt the BCD change point detection framework to analyze changes between two subject groups. We apply BCD-G to two groups, a group of smart home residents with MCI and/or dementia (CI) and a healthy, age-matched control group of smart home residents (HC). To explain the source of detected change, we analyze the activity and feature differences between the two groups.

A. Feature-based Change Analysis

The tree in Figure 3 shows the most discriminating features VC used to compute group-to-group CIΔHC. At the top of the tree is walking speed, which is commonly listed in the literature as a strong predictor of mortality [37]. Walking speed for the CI group averaged over all participants' days is 7.690 ± 6.243 ft/s, while walking speed for the HC group is 17.000 ± 9.832 ft/s. This large disparity explains why the first rule of the tree is walking speed ≤ 15.467 ft/s. Activities that are near the top of the tree include Bed to Toilet Transition, Dress, and Leave/Enter Home, implying healthy smart home residents tend to dress quicker, transition from the bed to the toilet quicker, and leave the home later in the day (see Figure 3 for the decision tree).

There are additional features that produce large differences between the two groups but are not necessarily discriminatory enough to be included near the top of the decision tree. These include activity duration variance, activity entropy, daytime sleep duration, and nighttime sleep duration. The cognitively impaired group has greater variance in the duration of the activities they perform (CI 9642.894 seconds; HC 8985.282 seconds), and in the entropy of those activities (CI 1.745 ± 0.291; HC 1.579 ± 0.473). The CI participants sleep more than the HC participants, with a higher daytime sleep duration (CI 1.348 ± 1.357 hours; HC 1.001 ± 1.350 hours) and nighttime sleep duration (CI 4.777 ± 2.125 hours; HC 4.129 hours ± 2.450 hours). Features that do not seem to differ much between the two groups include the number of unique activities (CI 13.724 activities per day; HC 13.657 activities per day), the number of nighttime sleep interruptions (CI 2.629 per night; HC 2.524 per night), and circadian rhythm strength (CI 0.745; HC 0.717). This last result is particularly interesting because of how informative the CRS values have been found to be in similar research comparing cognitively healthy and cognitively impaired groups. Paavilainen *et al.* found a CRS of ~0.3 was common for nondemented nursing home residents, while a CRS of ~0.5 was common for demented nursing home residents [38]. The CI and HC groups both have CRS values above 0.5,



a) PCAR group change score distributions.

b) SEP group change score distributions.

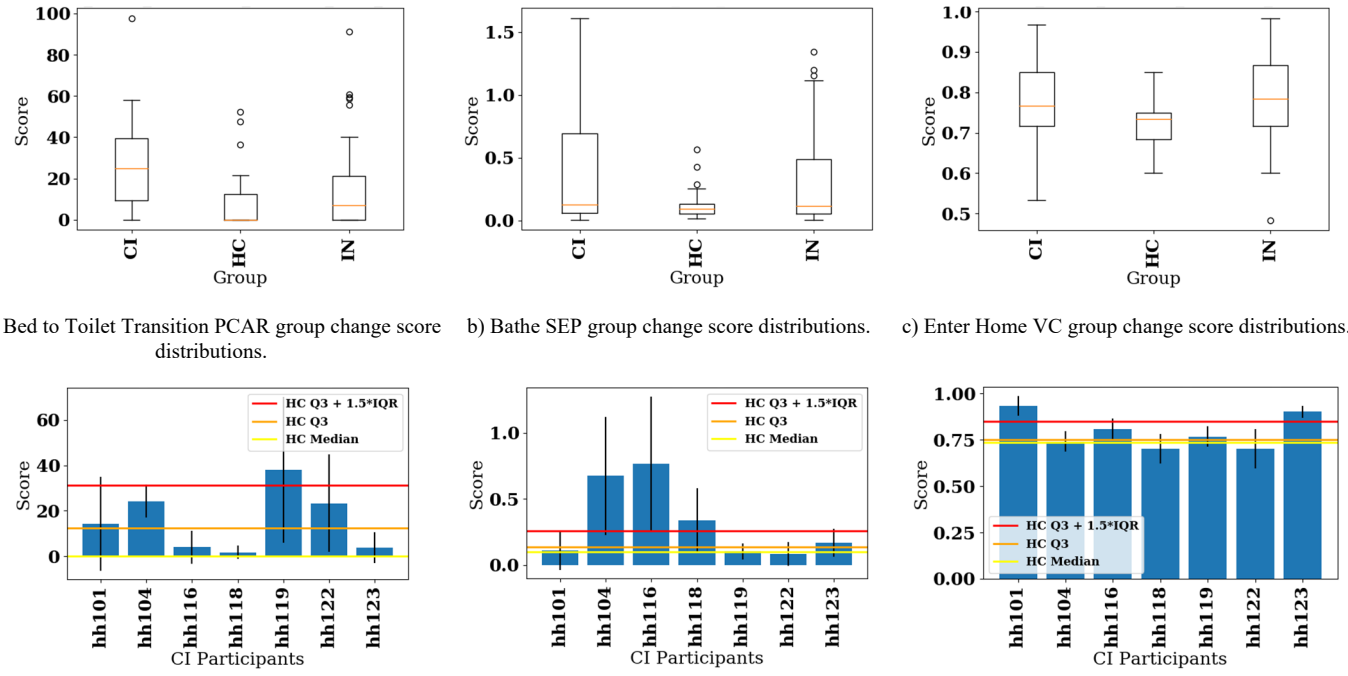
c) VC group change score distributions.

d) PCAR change scores for individual CI participants.

e) SEP change scores for individual CI participants.

f) VC change scores for individual CI participants.

Fig. 4. CIΔHC pairwise change score results for PCAR, SEP, and VC. Top: Distribution of change scores shown as box plots. Bottom: Pairwise changes for each participant in the CI group shown as bar plots. Bars show mean pairwise change scores while error lines on the bars show +/- one standard deviation of pairwise change scores. Three different values of the control group scores are shown to provide context for significant changes. CI = cognitive impairment group, HC = healthy control group, IN = intergroup, Q3 = third quartile (75th percentile), IQR = interquartile range.



a) Bed to Toilet Transition PCAR group change score distributions.

b) Bathe SEP group change score distributions.

c) Enter Home VC group change score distributions.

d) Bed to Toilet Transition PCAR change scores for individual CI participants.

e) Bathe SEP change scores for individual CI participants.

f) Enter Home VC change scores for individual CI participants.

Fig. 5. CIΔHC pairwise change score results for the activities with distinct differences observed between the two groups. Top: Distribution of change scores shown as box plots. Bottom: Pairwise changes for each participant in the CI group shown as bar plots. Bars show mean pairwise change scores while error lines on the bars show +/- one standard deviation of pairwise change scores. Three different values of the control group scores are shown to provide context for significant changes. CI = cognitive impairment group, HC = healthy control group, IN = intergroup, Q3 = third quartile (75th percentile), IQR = interquartile range.

indicating a high variability of activity from night to previous day; however, our results are not directly comparable to Paavilainen *et al*. because different measurements are used to quantify "activity". Paavilainen *et al*. used activity counts from wearable actigraph sensors, while in this paper we used activity labels from smart home sensors.

B. Activity-based Change Analysis

Using BCD-G with VC, we can investigate the strength of difference between the CI and HC groups for each activity. Of the 16 activities, all but one activity has statistically significant detected changes between CI and HC. The Other Activity is a catch-all label for behavior in the home that does not fit one of the original 40 activities. Considering this, it seems logical that behavior classified as Other Activity is likely somewhat random, regardless of which group a participant belongs to. The activities with the largest VC-detected change scores include Dress (0.710), Leave Home (0.698), Enter Home (0.695), and Meds (0.695). With the exception of Meds (which is not included in the decision tree at all), these features are all near the top in the decision tree in Figure 3 that was generated using all activities.

Unlike the group-to-group comparison results which provide single change score values, the pair-to-pair comparison results provide a distribution of change scores for three different groups, CI, HC, and IN (CI and HC combined). These distributions are visualized in Figure 4 (all activities included in BCD-G) and Figure 5 (three samples of single activity BCD-G results). Figure 4 shows that using PCAR for change detection captures change between the two groups, with the CI exhibiting a higher median change score and lower spread (see Figure 4a). Of the three algorithms, PCAR detects the most CI individual change compared to the distribution of the HC group (see Figure 4b). This is likely due to PCAR analyzing the hourly activity probability time series directly, instead of using summary features values as input like SEP and VC. The boxplot distributions for SEP (see Figure 4c) and VC (Figure 4d) do not show as much difference as PCAR (see Figure 4a) between the groups. Comparing the bar charts across all three algorithms does not show the algorithms agree about which participants exhibit the most change. Because of this ambiguity, we next look at each individual activity to narrow in on the individual participant's greatest source of behavior change.

Of the 16 activities, we included box plots and bar plots in Figure 5 for Bed to Toilet Transition (PCAR scores), Bathe (SEP scores), and Enter Home (VC scores). The CI group exhibits a large amount of variability in their routines for these three activities compared to the HC group (see Figures 5a-c). At the individual participant level, 4/7 CI smart home residents exceeded the 75th percentile of HC group change for these three activities; however, this is not a consistent set of 4 participants across all three activities. Furthermore, a CI participant with change scores outside the HC 75th percentile for all three activities does not exist. To investigate this further, we found that on average the participants demonstrated behavior changes greater than the HC 75th percentile for about 3 activities (4.286, 3.143, and 3.000 activities for PCAR, SEP, and VC-detected changes, respectively). This suggests the CI participants exhibit radically different change from the healthy population on only a handful of activities at a time.

For commercial products focusing on extending independence in persons with cognitive impairment, our research suggests that change point detection for the clinically relevant activities of walking speed (from bed to toilet), bathing routines, and enter/exit home activities for CI persons should be included in smart home product design. Activities representing the milieu of symptoms associated with cognitive decline may require varied algorithmic approaches. Products detecting these changes can facilitate a type of "clinical triage" to support clinicians, caregivers, and patients. Early interventions are key to reducing premature transition to residential memory care and reducing CI-associated healthcare costs [39].

VI. Conclusion

Using daily smart home activity data and our Behavior Change Detection for Groups algorithmic framework, we are able to identify key behavior differences between groups of smart home residents. These differences include changes in duration, complexity, and pattern of common activities of daily living, such as Bed to Toilet Transition, Dress, Leave/Enter Home, and Meds. We also analyze changes in clinically relevant behavior features, such as quantifying how much slower cognitively impaired subjects walk in their home as compared to age-matched healthy subjects.

Limitations of this work include the small sample size of seven cognitively impaired smart home participants and seven age-matched smart home participants. Also, the expansion to BCD-G that performs pairwise comparisons does not scale due to its exponential time complexity. The approach is best for comparing small groups in an offline setting; however, for large groups and/or an online deployment of the algorithm, we can overcome this limitation by using a *k*-nearest neighbors approach to sampling which participants within a group are most important to compare with. In the future, we plan to move BCD-G to an online setting to support clinician-in-the-loop and preventative healthcare. With BCD-G, we are able to compare the in-home behavior of a single participant to a cognitively impaired subgroup and a healthy control subgroup to identify similarities and differences. With this information, clinicians can remotely determine at-risk behavior and provide timelier and more precise medical and social interventions.

Acknowledgment

The authors would like to thank Maureen Schmitter-Edgecombe and Beiyu Lin for their assistance with this study.

References

[1] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowl. Inf. Syst.*, vol. 51, no. 2, pp. 339–367, May 2017, doi: 10.1007/s10115-016-0987-z.

[2] C. Galambos, M. Skubic, S. Wang, and M. Rantz, "Management of dementia and depression utilizing in-home passive sensor data," *Gerontechnology*, vol. 11, no. 3, pp. 457–468, Jan. 2013, doi: 10.4017/gt.2013.11.3.004.00.

[3] B. Fadem, *Behavioral Science in Medicine*. Lippincott Williams & Wilkins, 2012.
[4] A. R. Roberts, K. B. Adams, and C. B. Warner, "Effects of chronic illness on daily life and barriers to self-care for older women: A mixed-methods exploration," *J. Women Aging*, vol. 29, no. 2, pp. 126–136, Mar. 2017, doi: 10.1080/08952841.2015.1080539.
[5] J. M. Orient and J. D. Sapira, *Sapira's Art & Science of Bedside Diagnosis*. Lippincott Williams & Wilkins, 2010.
[6] G. Sprint, D. J. Cook, R. Fritz, and M. Schmitter-Edgecombe, "Using Smart Homes to Detect and Analyze Health Events," *Computer*, vol. 49, no. 11, pp. 29–37, Nov. 2016, doi: 10.1109/MC.2016.338.
[7] G. Sprint, D. Cook, R. Fritz, and M. Schmitter-Edgecombe, "Detecting Health and Behavior Change by Analyzing Smart Home Sensor Data," in *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2016, pp. 1–3, doi: 10.1109/SMARTCOMP.2016.7501687.
[8] D. C. Mohr, M. Zhang, and S. M. Schueller, "Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning," *Annu. Rev. Clin. Psychol.*, vol. 13, no. 1, pp. 23–47, 2017, doi: 10.1146/annurev-clinpsy-032816-044949.
[9] M. V. Vitiello, "Sleep in Normal Aging," *Sleep Med. Clin.*, vol. 1, no. 2, pp. 171–176, Jun. 2006, doi: 10.1016/j.jsmc.2006.04.007.
[10] J. C. Ayena, L. D. Chapwouo T., M. J.-D. Otis, and B.-A. J. Menelas, "An efficient home-based risk of falling assessment test based on Smartphone and instrumented insole," in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, 2015, pp. 416–421, doi: 10.1109/MeMeA.2015.7145239.
[11] M. Mancini *et al.*, "Continuous Monitoring of Turning Mobility and Its Association to Falls and Cognitive Function: A Pilot Study," *J. Gerontol. Ser. A*, vol. 71, no. 8, pp. 1102–1108, Aug. 2016, doi: 10.1093/gerona/glw019.
[12] M. O'Brien, C. Mummidisetty, X. Bo, and C. Poellabauer, "Quantifying community mobility after stroke using mobile phone technology," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 161–164.
[13] M. Boukhechba, Y. Huang, P. Chow, K. Fua, B. Teachman, and L. Barnes, "Monitoring social anxiety from mobility and communication patterns," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 749–753.
[14] M. A. U. Alam, N. Roy, A. Gangopadhyay, and E. Galik, "A smart segmentation technique towards improved infrequent non-speech gestural activity recognition model," *Pervasive Mob. Comput.*, vol. 34, pp. 25–45, Jan. 2017, doi: 10.1016/j.pmcj.2016.06.015.
[15] D. J. Cook, M. Schmitter-Edgecombe, and P. Dawadi, "Analyzing Activity Behavior and Movement in a Naturalistic Environment Using Smart Home Techniques," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 6, pp. 1882–1892, Nov. 2015, doi: 10.1109/JBHI.2015.2461659.
[16] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated Cognitive Health Assessment Using Smart Home Monitoring of Complex Tasks," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 43, no. 6, pp. 1302–1313, Nov. 2013, doi: 10.1109/TSMC.2013.2252338.
[17] A. Nait Aicha, G. Englebienne, and B. Kröse, "Continuous Gait Velocity Analysis Using Ambient Sensors in a Smart Home," in *Ambient Intelligence*, Cham, 2015, pp. 219–235, doi: 10.1007/978-3-319-26005-1_15.
[18] D. Austin, T. L. Hayes, J. Kaye, N. Mattek, and M. Pavel, "Unobtrusive monitoring of the longitudinal evolution of in-home gait velocity data with applications to elder care," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 6495–6498.
[19] N. D. Darnall *et al.*, "Application of machine learning and numerical analysis to classify tremor in patients affected with essential tremor or Parkinson's disease," *Gerontechnology*, vol. 10, no. 4, pp. 208–219, Mar. 2012, doi: 10.4017/gt.2012.10.4.002.00.
[20] A. Alberdi *et al.*, "Smart Home-Based Prediction of Multidomain Symptoms Related to Alzheimer's Disease," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1720–1731, Nov. 2018, doi: 10.1109/JBHI.2018.2798062.
[21] S. Hellmers *et al.*, "Towards a minimized unsupervised technical assessment of physical performance in domestic environments," in *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2017, pp. 207–216.
[22] A. Akl, J. Snoek, and A. Mihailidis, "Unobtrusive Detection of Mild Cognitive Impairment in Older Adults Through Home Monitoring," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 2, pp. 339–348, Mar. 2017, doi: 10.1109/JBHI.2015.2512273.
[23] J. Petersen, S. Thielke, D. Austin, and J. Kaye, "Phone behaviour and its relationship to loneliness in older adults," *Aging Ment. Health*, vol. 20, no. 10, pp. 1084–1091, Oct. 2016, doi: 10.1080/13607863.2015.1060947.
[24] S. Robben, G. Englebienne, and B. Kröse, "Delta Features From Ambient Sensor Data are Good Predictors of Change in Functional Health," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 986–993, Jul. 2017, doi: 10.1109/JBHI.2016.2593980.
[25] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: A Smart Home in a Box," *Computer*, vol. 46, no. 7, pp. 62–69, Jul. 2013, doi: 10.1109/MC.2012.328.
[26] N. C. Krishnan and D. J. Cook, "Activity Recognition on Streaming Sensor Data," *Pervasive Mob. Comput.*, vol. 10, pp. 138–154, Feb. 2014, doi: 10.1016/j.pmcj.2012.07.003.
[27] J. Merilahti, P. Viramo, and I. Korhonen, "Wearable Monitoring of Physical Functioning and Disability Changes, Circadian Rhythms and Sleep Patterns in Nursing Home Residents," *IEEE J. Biomed. Health Inform.*, vol. PP, no. 99, pp. 1–1, 2015, doi: 10.1109/JBHI.2015.2420680.
[28] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-Point Detection in Time-Series Data by Relative Density-Ratio Estimation," *Neural Netw.*, vol. 43, pp. 72–83, Jul. 2013, doi: 10.1016/j.neunet.2013.01.012.
[29] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Modeling Patterns of Activities Using Activity Curves," *Pervasive Mob. Comput.*, 2015, doi: 10.1016/j.pmcj.2015.09.007.
[30] G. Sprint, D. J. Cook, and M. Schmitter-Edgecombe, "Unsupervised detection and analysis of changes in everyday physical activity data," *J. Biomed. Inform.*, vol. 63, pp. 54–65, Oct. 2016, doi: 10.1016/j.jbi.2016.07.020.
[31] S. Wang, M. Skubic, and Y. Zhu, "Activity Density Map Visualization and Dissimilarity Comparison for Eldercare Monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 607–614, Jul. 2012, doi: 10.1109/TITB.2012.2196439.
[32] T.-H. Tan *et al.*, "Indoor Activity Monitoring System for Elderly Using RFID and Fitbit Flex Wristband," in *2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2014, pp. 41–44, doi: 10.1109/BHI.2014.6864299.
[33] S. Hido, T. Ide, H. Kashima, H. Kubo, and H. Matsuzawa, "Unsupervised Change Analysis Using Supervised Learning," in *Advances in Knowledge Discovery and Data Mining*, Springer Berlin Heidelberg, 2008, pp. 148–159.
[34] S. Aminikhanghahi, T. Wang, and D. J. Cook, "Real-Time Change Point Detection with Application to Smart Home Time Series Data," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 5, pp. 1010–1023, May 2019, doi: 10.1109/TKDE.2018.2850347.
[35] S. Aminikhanghahi and D. J. Cook, "Enhancing activity recognition using CPD-based activity segmentation," *Pervasive Mob. Comput.*, vol. 53, pp. 75–89, Feb. 2019, doi: 10.1016/j.pmcj.2019.01.004.
[36] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc.*, vol. 57, no. 1, pp. 289–300, 1995.
[37] S. Studenski *et al.*, "Gait Speed and Survival in Older Adults," *JAMA*, vol. 305, no. 1, pp. 50–58, Jan. 2011, doi: 10.1001/jama.2010.1923.
[38] P. Paavilainen *et al.*, "Circadian Activity Rhythm in Demented and Non-Demented Nursing-Home Residents Measured by Telemetric Actigraphy," *J. Sleep Res.*, vol. 14, no. 1, pp. 61–68, Mar. 2005, doi: 10.1111/j.1365-2869.2004.00433.x.
[39] A. Shapiro and M. Taylor, "Effects of a Community-Based Early Intervention Program on the Subjective Well-Being, Institutionalization, and Mortality of Low-Income Elders | The Gerontologist | Oxford Academic," *The Gerontologist*, vol. 42, no. 3, pp. 334–341, Jun. 2002.