



Published in final edited form as:

ACM Trans Comput Healthc. 2022 July ; 3(3): . doi:10.1145/3508020.

Detecting Smartwatch-based Behavior Change in Response to a Multi-domain Brain Health Intervention

DIANE J. COOK*,
MIRANDA STRICKLAND,
MAUREEN SCHMITTER-EDGECOMBE

Washington State University

Abstract

In this study, we introduce and validate a computational method to detect lifestyle change that occurs in response to a multi-domain healthy brain aging intervention. To detect behavior change, digital behavior markers (DM) are extracted from smartwatch sensor data and a Permutation-based Change Detection (PCD) algorithm quantifies the change in marker-based behavior from a pre-intervention, one-week baseline. To validate the method, we verify that changes are successfully detected from synthetic data with known pattern differences. Next, we employ this method to detect overall behavior change for $n=28$ BHI subjects and $n=17$ age-matched control subjects. For these individuals, we observe a monotonic increase in behavior change from the baseline week with a slope of 0.7460 for the intervention group and a slope of 0.0230 for the control group. Finally, we utilize a random forest algorithm to perform leave-one-subject-out prediction of intervention versus control subjects based on digital marker delta values. The random forest predicts whether the subject is in the intervention or control group with an accuracy of 0.87. This work has implications for capturing objective, continuous data to inform our understanding of intervention adoption and impact.

Keywords

Behavior change detection; Machine learning from time series; Activity recognition; Behavior intervention

1 INTRODUCTION

The connection between behavior and health is undeniable. Unhealthy behaviors and habits are implicated in as much as 40% of premature deaths, in addition to their unfavorable effects on health disparities, quality of life, and economics [1]. Advances in sensor-driven technologies for behavior monitoring suggest that these technologies can monitor intervention adherence and impact in the wild [2]-[4]. However, our current understanding of intervention on a person's entire lifestyle is limited because studies have been based on static "snapshots" or assessment of a specific set of behavior parameters (e.g. sleep

*School of Electrical Engineering and Computer Science. djcook@wsu.edu.

quality, activity level), rather than data-driven analysis of overall behavior change. Given this limitation, there is a need to introduce algorithmic methods for quantifying behavior change that is characterized by a person's entire pattern of behavior.

In this work, we introduce an automated method to perform continuous monitoring and change detection of daily activities, in naturalistic settings, using sensor-based observation and machine learning-based data analysis. To do this, we perform the steps shown in Figure 1. We collect wearable sensor data for two months, while subjects perform their daily routines. We process the collected sensor data by merging data from multiple watches, imputing missing values, and automatically associating data sequences with corresponding activity labels. Next, we extract digital behavior markers from activity-labeled data that summarize the key features of each person's behavior on an hourly and daily basis. Finally, we apply a behavior change detection to the digital behavior markers. This algorithm quantifies and describes behavior differences between time points.

To validate our approach, we analyze data from the active condition of a pilot multi-domain Brain Health Intervention (BHI) for midlife and older adults. We also compare the data with an age-matched control group that did not receive an intervention. The seven-week pilot intervention, called B-Fit, was designed to improve engagement in aging brain health behaviors by providing brain health psychoeducation combined with individual behavior goal setting, group problem solving, and social support [5]. Participants wore smartwatches during the intervention that tracked behaviors. Control participants similarly wore smartwatches. Seven risk factors were targeted by the intervention: nutrition, stress, social engagement, cognitive engagement, cardiovascular risk factors, physical activity, and sleep. To promote new patterns of engagement in healthy brain aging behaviors and better sustain behavior change, participants set individualized, intrinsically motivating, and manageable new goals (5-10 minute) each week (e.g., take stairs rather than elevator). Results from the primary analysis revealed that self-reported engagement in healthy brain aging behaviors was higher for the B-Fit condition and an education-only condition at post-intervention, while there was no difference for the waitlist condition. Furthermore, after adjusting for baseline brain health behaviors, compared to the waitlist condition, participants in the B-Fit condition self-reported greater engagement in healthy brain aging behaviors post-intervention [5].

Because BHI is a multi-domain intervention where participants set their own diverse goals, computational methods are needed to compliment self-report data by providing a method to objectively detect and quantify change in overall behavior patterns, rather than analyzing a small set of target markers. We initially apply our proposed computational approach to synthetic data to validate their ability to detect known embedded changes in behavior data. We then apply our methods to smartwatch data collected from BHI and control subjects. We hypothesize that behavior change over time will be exhibited among BHI subjects. Because change can be attributed to natural behavior variability, we further postulate that behavior change will be larger for BHI subjects than for control subjects. In addition to quantifying change over time for the participants, we also employ supervised learning to automatically classify each subject as belonging to the intervention or control group, based solely on

changes in the extracted digital markers between baseline and the end of the data collection period.

This article makes the following contributions: (1) We offer a novel method to extract digital markers from time series sensor data that characterize a person's activity routine. (2) We introduce a way to analyze behavior across time based on a permutation detection algorithm. (3) We computationally analyze the impact of a multi-domain brain intervention study, contrasting behavior change of intervention participants with an age-matched control group.

2 RELATED WORK

2.1 Activity Learning

To characterize a person's overall behavior routine, we extract digital behavior markers from sensor data that are automatically labeled with corresponding activity categories. Performing human activity recognition from wearable sensors has become a popular topic for researchers to investigate [6], [7]. Because we can collect uninterrupted data in all locations each person visits, wearable sensors are a natural choice for this investigation. Earlier approaches to activity recognition consider many diverse learning models, including decision trees, nearest neighbors, clusters, and classifier ensembles [8]-[10]. With the rising popularity of deep learning, these methods have also been explored for their ability to learn deep features that are useful in expressing sensed activities [11]-[14]. The field has matured to the point that numerous public datasets are now available for comparative analysis of recognition methods [11].

Limitations with many of these existing methods is that they focus on basic, repetitive movement types and are typically evaluated in laboratory settings. Recognizing postures and ambulatory motions such as sitting, standing, climbing, lying down, and running are useful for monitoring gait characteristics and activity levels. However, we are interested in determining change to a person's entire behavioral routine. This is frequently characterized by more complex patterns, such as basic and instrumental activities of daily living (iADLs) [15], [16]. A few researchers consider activity categories that combine ambulation with context (e.g., standing indoors versus outdoors) [8], [17], [18]. They accomplish this by adding location information. This location information is provided by ambient motion sensors inside a specific physical environment [17] or a grid location within a specific physical region [8]. A challenge with analyzing specific locations is that the insights do not generalize well to new individuals. Lin and Hsu [19] introduced generalizable location features that we employ in our method, including heading change rate, distance covered, and velocity change rate. We further integrate ideas from Boukhechba et al. [20], who cluster location readings to detect person-specific frequented spots. While we extend these previous methods to perform real-time activity labeling of smartwatch sensor data in naturalistic settings, the success these researchers have documented in recognizing a wide assortment of activities suggests the general applicability of activity learning and labeling for smartwatch data.

2.2 Digital Behavior Markers

Sensor data offer substantial insights on a person's behavior as well as their health. Now that wearable sensors are decreasing in cost and are deployable in real-world settings, researchers have discovered that they can link behavior interventions with specific, sensor-observed target behavior outcomes. As wearable sensor systems also become more ubiquitous, researchers have increasingly used these platforms to impact and measure change in human behavior. Typically, these prior works target a specific behavior. Popular target behaviors include activity levels [21], [22] and sleep [23]. In these cases, change can be quantified by a single variable such as step counts, sleep duration, or a clinical score [24]. Jain et al. designed a method to analyze the univariate change in the state of a person or place, such as the amount of activity in a region of the home [25]. Wang et al. calculated the variability of a single variable, then joined the variability of multiple features into a single vector to predict personality traits [26]. In a method that utilizes mobile platforms for intervention delivery as well as evaluation, Costa et al. [27] delivered haptic feedback to participants intended to slow their heart beats while completing math tests. They observed impacts based on self-reported anxiety, sensor-measured heart rate variability, and test performance. In a study conducted by Stanley et al. [28], individuals with dementia received coaching in behavior techniques designed to reduce anxiety. Here, outcomes were measured through clinician-rated and collateral report clinical scores targeted at assessing anxiety. Additional studies monitor single wearable markers to assess intervention impact for target conditions such as substance use disorder [29] and diabetes management [30].

In other work, researchers targeted specific sensor-observed behavior markers as a mechanism for assessing the relationship between lifestyle and health. Specifically, Dhana et al. [31] quantify healthy behavior as a combination of nonsmoking, physical activity, alcohol consumption, nutrition, and cognitive activities. Individuals who scored higher on this behavior metric had lower risk of Alzheimer dementia. Other researchers have also found that sensor-based behavior patterns are predictive of cognitive health [32],[33]. Li et al. [33] found that physical activity was predictive of Alzheimer's disease, while Aramendi et al. [32] were able to predict cognitive health and mobility from activity-labeled sensor data.

These studies provide evidence that wearable sensors afford the ability to monitor intervention impact and to assess a person's cognitive health. Within this area of investigation, our proposed approach is unique because we investigate a computational method to analyze intervention impact on the pattern of a person's activity-based routine. This holistic approach to sensor analysis of behavior is motivated by the design of a holistic, customized, sustainable healthy behavior intervention.

2.3 Time Series Change Detection

A critical component of this work is automatically detecting and quantifying change in behavior data observed longitudinally. Multiple options are available for computing change between two windows of multivariate time series data [34]. Supervised methods can classify a pair of time windows as "no change" or "change". Such methods can also identify changes of a specific nature (e.g., a change from sedentary to active behavior),

but these algorithms must be supplied with a sufficient amount of labeled training data [35]. Unsupervised approaches have been introduced as well. While some methods rely on pre-designed parametric models [36], [37], others introduce more flexible non-parametric variations by directly estimating the ratio of probability densities for the two samples. In these cases, a larger ratio implies a greater likelihood that a change occurred between the two samples. Methods that fit in this category are KLIEP which employs Kullback-Leibler (KL) divergence as a ratio estimator [38], uLSIF which employs Pearson divergence [38], and Relative uLSIF (RuLSIF) [38]. In this work, we utilize a permutation-based change detection method. This method is not only suited to comparing larger windows of sensor data but also incorporates a novel permutation mechanism to determine the significance of change over time.

2.4 Brain Health Intervention

An accumulating body of research suggests that adopting preventative health behaviors may promote healthy brain aging and slow cognitive decline [39]. Given that the etiology of dementia is heterogeneous and influenced by multiple risk factors, several recent large scale prevention trials have focused on evaluation of multi-domain interventions [40], [41]. However, these interventions are currently limited by expensive treatment regimens and highly prescriptive goals that are difficult to sustain post-intervention. Furthermore, research has found that intervention success is promoted by healthy behavior education as well as designing interventions that are transferable out of the lab and into daily lives [42], [43]. In contrast with prior methods, the pilot group brain health intervention involved creating individualized behavior change goals across multiple risk factor domains that could be integrated into the everyday lifestyles of participants and more easily sustained over time.

3 BRAIN HEALTH INTERVENTION STUDY

The objective of the pilot brain health intervention study was to develop and evaluate an intervention designed to promote changes in healthy brain-behavior for midlife and older adults that is practical and easily replicated. The rationale is that a preventative brain health intervention designed to reduce disability in older age will be most successful if it meets individuals where they are and promotes healthy behavior change goals that can be more easily sustained. Furthermore, the approach is motivated by findings that multiple risk factors influence the development of dementia. Researchers estimate that almost half of dementia cases are attributable to modifiable risk factors including physical inactivity and depression [44]. Studies [45] indicate that engaging in healthy lifestyles behaviors across multiple domains including physical activity, cognitive engagement, and stress management can slow cognitive decline [46]-[51] and promote brain neuroplasticity [52], [53]. There currently exists a gap in the design of methods to assist midlife and older adults in learning to incorporate such healthy behaviors into their everyday lives in a way that leads to sustained, holistic behavior change.

3.1 Intervention Design

Our brain health intervention, called B-Fit, is a seven-week intervention focused on providing brain health psychoeducation combined with individual behavior goal setting,

group problem solving, and social support. The primary pilot clinical intervention study included 3 conditions: B-Fit, education-only, and waitlist [5]. Of the 68 participants allocated to the B-fit intervention in the primary study, 28 of the participants wore a smart watch during the intervention. These individuals were members of three separate B-fit groups. The B-fit intervention was delivered by two clinicians with groups of 8-13 participants. One learning topic was introduced to the group each week. During didactics, information about each topic was discussed with an emphasis on empirical research describing mechanisms by which lifestyles changes can improve brain health (e.g., reduce blood pressure and oxidative stress, increase cerebral blood flow) and contribute to a healthy aging phenotype. This approach is based on literature indicating the importance of linking healthy lifestyle behaviors with brain health and prevention research [54], [55]. The goal of this paper is to introduce and validate a mechanism for assessing broad-spectrum lifestyle changes associated with this individualized, multi-domain intervention.

Table 1 lists the primary health topics covered during the weekly sessions. Each week, following the educational material presentation, participants developed a plan to integrate a personalized, realistic, sustainable, and positive lifestyle change into their everyday life (e.g., take a 10-minute walk daily after lunch). Health goals were built across multiple lifestyle domains consistent with the education topics (e.g., diet, socialization, sleep). Because a new goal was introduced each week, participants were assisted with developing goals for each topic that could be completed in about 5-10 minutes. At the beginning of each session, for each defined goal, participants indicate whether they made no change (0), partly met (1), completely met (2), or exceeded (3) their goal(s). 26 of the 28 participants that wore a smartwatch regularly reported these values. The average self-reported adherence over the 8-week intervention was 1.7 (range [0.5, 2.3], standard deviation 0.44).

3.2 Data Collection

In addition to evaluating self-reported outcomes, determining the impact of interventions such as B-Fit on the pattern of behaviors can be implemented by passively and continuously monitoring behavior across time. Automated approaches to such ecologically-valid monitoring and assessment pave the way for designing more effective treatment and quantitatively measuring the outcome of pharmaceutical and behavioral interventions. This ability will be particularly valuable in rural locations where access to the clinic or lab is limited [56]. Ecologically-valid monitoring implies that a person's routine behavior is observed in naturalistic settings, rather than asking individuals to perform scripted activities in controlled settings. This is particularly important for the B-Fit intervention because we want to assess the impact of the intervention on a person's entire set of activities, not on a few selected markers. Currently, holistically assessing change in overall behavior patterns in such a manner is difficult outside of manually observing a person's daily routine. We determine whether behavior change exists using a computational behavior change detection method. We further evaluate the pattern of behavior change over the course of the intervention.

To provide objective behavior routine and change measures, we gave 28 of the B-Fit participants two Apple Watches to wear continuously, one during the day and one at night

(one watch was charged while the other was worn). We collected timestamped readings at 10Hz for the following sensors: 3d acceleration, 3d rotation and rotation rate, course, speed, and 3d location. Participants received two motivational quotes through the smartwatch at random times during the day to support their lifestyle changes. Data were continuously collected at baseline (before intervention) and each week of the intervention. All data were collected before quarantine restrictions due to COVID-19.

3.3 Participants

All study participants are community-dwelling adults. We recruited participants from eastern Washington and northern Idaho. Exclusion criteria included a clinical diagnosis of dementia, inability to provide own consent, unstable or severely disabling disease (e.g., organ failure), and inability to complete assessment and intervention protocols due to communication, vision, hearing, or other medical difficulties. Prior to participation, participants were screened by phone with a medical/health interview and the Telephone Interview of Cognitive Status (TICS, [57]). Participants were excluded if they fell in the Impaired range on the TICS (score < 26) or met study exclusion criteria. The study was approved by the Washington State University Institutional Review Board.

We analyze data for 28 study participants who received the active brain health education and goal-setting intervention (B-Fit) described in Section 2.1 and continuously wore smartwatches. This group is labeled brain health intervention, or ‘BHI’. To provide a comparison (non-intervention) group for behavior change detection, we analyze continuous smartwatch sensor data collected for 17 additional community-dwelling older adults who did not differ in age, $t(43) = 1.08$ $p = .28$, education, $t(43) = .17$ $p = .87$, or sex, $\chi^2(1, N = 45) = .33$, $p = .57$, from the BHI group. This group is labeled ‘NI’. These participants were older adults who transformed their homes into smart homes and participated in a non-intervention, monitoring study designed to help validate smart home health algorithms. Table 2 summarizes the demographics for these participants.

4 PROCESSING COLLECTED DATA

We collected multiple weeks of smartwatch sensor data for each of the subjects, totaling 28,615,680 sets of readings for BHI participants and 18,464,560 sets of readings for the non-intervention group. We examine the first nine weeks of data collection for each subject. This length of time corresponds to the one week of baseline, seven weeks of intervention and one post intervention week for the BHI participants. Our goal is to examine change in behavior that occurs over the course of the intervention. Before analyzing the data, we first merge data from multiple watches, impute missing data, and label collected data with activity labels.

4.1 Data Preparation

Participants were given a white watch to wear during the day and a black watch to wear at night, affording time to charge the devices. The smartwatches indicate when they are on and off the charging station, allowing us to determine when data should be analyzed from one device or the other. However, participants were not consistent in wearing the designed

watches during the specified times. To ensure we analyze the appropriate data, we examine the acceleration and rotation-based movement throughout each day. We merge day and night data collections into one by identifying “swapping” periods, or time points at which the subject swapped one watch for another. This was defined as a change point in the time series when one watch was at rest (total movement below a threshold amount) and the other was active. We create one merged set of sensor readings for each day by incorporating data at each time unit from the active watch.

Additionally, gaps occurred during data collection. Data for participants were 94.24% complete. Gaps in the data collection occurred during times when subjects forgot to charge and wear the watch. In these cases, we impute missing feature values based on the median value for the particular sensor given the day of the week, time of day, and neighboring context (readings before and after the gap).

4.2 Activity Labeling

We automatically label each set of sensor readings with a corresponding activity name. Activity labeling offers several benefits for our analysis. First, activities provide a vocabulary for expressing human behavior. From activity-labeled sensor data we can see how a person spends their day and observe changes in these activity routines over time. Second, given activity-labeled data we can target specific activity categories, such as work, exercise, and sleep, to ascertain whether the amount of time that is spent on these specific activities is changing, if the location where they occur changes, and if they occur at regular times.

We label sensor data using an activity recognition algorithm. Activity recognition maps a sequence of sensor readings onto an activity label. A time series data stream is an infinite sequence of elements $S = \{x_1, \dots, x_i, \dots\}$, where x_i is a d-dimensional data vector arriving at time stamp i . Given a set of predefined activity categories, A , an activity recognition algorithm maps a sequence of time series data points $\{x_m, \dots, x_k, \dots, x_n\}$, observed as part of time series S , onto an activity label $a_j \in A$. We map a fixed-length sequence of elements onto the activity label using the process illustrated in Figure 2.

To ensure that we label and monitor complex activities as well as basic movement-based activities, we also incorporate location information. Because we want to ensure that the model generalizes to multiple persons, we do not utilize specific location values. Instead, we construct person-invariant location features. This approach supports generalization without needing to employ techniques such as domain adaptation, that requires a large number of labeled training instances for the initial group of subjects [58]-[60]. One such person-invariant location feature is the location type, obtained through reverse geocoding using an OpenStreetMap. We group these location categories into residence, workplace, road, and other. Other location features are calculated for each sensor sequence and include the number of changes in orientation in the sequence, number of stops/starts, the sequence overall trajectory, and the distance traversed.

Attempting to perform real-time human activity recognition “in the wild” in naturalistic settings raises numerous challenges. One such challenge is collecting ground-truth labeled

data. Users can be queried for their activities. However, these queries need to be answered in-the-moment to avoid retrospective error. While active learning and semi-supervised learning can reduce the number of queries [61], [62], the resulting interactions will interrupt many of the activities we wish to monitor for our brain health study. To make the intervention as naturalistic as possible, we do not ask the subjects to provide labels for their daily activities. Instead, we train an activity recognition algorithm using past data collected from 388 different smartwatch wearers. These earlier volunteers provided labels for their current activity when prompted at random times. These prompts resulted in 59,610 labeled instances for 12 activity categories. These activity categories are reflective of routine behavior and include: *chores, eat, entertainment, errands, exercise, hobby, hygiene, relax, school, sleep, travel, and work*.

Each five-second sequence of sensor data is considered as a separate instance. While shorter sequences have been used with success for ambulation recognition [63], we found that slightly longer durations are needed to capture context for recognizing complex activities in naturalistic settings. A feature vector is extracted from each five-second sensor sequence. The feature vector consists of standard statistical operations applied to each of the sensor types [64]. The vector also contains person-invariant location features. We experimented with multiple types of supervised learning algorithms and found random forest to perform the best for these data. The random forest also generates a prediction for one sequence before the next sensor reading arrives, allowing the activity recognition to occur in real time. The random forest employs bootstrapping and is configured to contain 100 trees with a depth limit of 10. Because the training data are not uniformly distributed across the activity classes, we weight instances inversely proportionate to the relative size of the corresponding activity class. Applying 3-fold cross validation to our labeled data, activity recognition yielded an accuracy of 0.86 for the 12 activity classes. For the remainder of this paper, we analyze sensor data that is labeled using a random forest model trained on all the labeled instances.

4.3 Extracting Digital Behavior Markers

Once continuous wearable sensor data are collected and labeled with activity classes, we extract digital markers that describe the person's behavior patterns. Extracting features from longitudinal sensor data is popular for analyzing specific behavior parameters. As mentioned in Section 2, researchers typically extract indicators that are specific to a behavior of interest. Depending on the goal of the study they may include amount of screen time, activity level, steps, incoming/outgoing phone calls and texts, number of steps, amount of screen time, and sleep. We are interested in characterizing a person's behavior routine. This is expressed as a set of digital markers that express how a person's time is spent on different activities, the locations a person frequents and the amount of time they spend there, and the traditional measures of overall movement types and durations. Obtaining digital behavior markers that express activity routines as well as direct sensor measurements of location and behavior are possible because we automatically label sensor sequences with activity labels, as described in Section 4.2. Thus, in addition to generating statistical summaries of raw sensor readings as listed in Figure 2 [65]-[68] we also extract markers that indicate how each

subject divides their time among the activity categories and where they spend their time on a routine basis.

Table 3 summarizes all of these digital behavior markers that we extract and analyze. Our activity recognition algorithm applies statistical operators (listed in Figure 2) to raw sensor readings, yielding a feature vector that is mapped to an activity label. When assembling digital behavior markers, we apply these same statistical operators to aggregated daily and hourly sensor values, activity information, and location information. Additionally, we compute a regularity index to each variable as defined by Wang et al. [26]. The regularity index determines how repetitive a person's behavior routine is. This is computed as the difference between hourly statistics occurring during the same hours across two different days. To calculate this index, data are first scaled to the range $[-0.5, 0.5]$. The regularity index comparing days a and b is then defined as shown in Equation 1.

$$RI_{a,b} = \sum_{t=1}^T \frac{x_t^a x_t^b}{T} \quad (1)$$

For our analysis, T is 24 hours and x_t^a represents the value of a behavior marker for hour t of day a . We average the regularity index over all pairs of days occurring within the same week, all pairs of weekdays occurring within the same week, and between different weeks for the same day-of-the-week value. Finally, we compute the diurnal and circadian rhythms for each of the hourly values. We follow guidelines by Depner et al. [69] to define circadian phase as constancy in the behavior routine. We quantify this definition by generating a periodogram for the marker values and calculating the normalized periodogram values for a 24-hour cycle (circadian) and for the 12 hours following the end of nighttime sleep (diurnal). Based on Fourier analysis, the spectral energy of each cycle frequency can be calculated as shown in Equation 2.

$$R = \left(\frac{2}{N} \sum_{i=1}^N x_i \cos\left(\frac{2\pi j t_i}{N}\right) \right)^2 + \left(\frac{2}{N} \sum_{i=1}^N x_i \sin\left(\frac{2\pi j t_i}{N}\right) \right)^2 \quad (2)$$

In Equation 2, N represents the number of samples in the time series, x_i represents the value of the measured variable at time i , and the spectral energy is computed for a particular frequency, j . We calculate the spectral energy for a range of possible frequencies. The resulting periodogram generates values for a set of possible cycle lengths. The corresponding circadian rhythm parameter is then represented by the normalized periodogram-derived value for a 24-hour cycle, while the 12-hour post-sleep cycle is computed for the diurnal parameter. At the completion of the feature extraction process, our system has constructed a vector containing 102 behavior features for each observed day, 40 features for each observed hour, and 863 overall features for each person.

5 DETECTING CHANGES IN ACTIVITY PATTERNS

We next consider the challenge of detecting and quantifying change in behavioral routine patterns. To address this problem, we introduce a method called Permutation-based Change

Detection, or PCD, that determines if a change exists between two windows of activity-labeled multivariate time series data. PCD quantifies the amount of change and determines the significance of the change. Algorithm 1 outlines the process. Let X denote a sample of multivariate time series data segmented into days. Each day, D , is further segmented into equal-sized time intervals, $D = \{x_1, x_2, \dots, x_m\}$. We let W denote a window of n days such that $W \subseteq X$.

We compare windows of data within time series X . Let W_i refer to a window that starts at day i . Then $W_i = X[i:i+n-1] = \{D_i, D_{i+1}, \dots, D_{i+n-1}\}$. The compared windows may represent consecutive time periods (e.g., days, weeks, months), or overlapping windows. In our case, we compare a single baseline window with each subsequent time period. Specifically, we compare the first week of observed behavior data, W_1 , with each following week $W_i, i \in \{2..9\}$. In our analysis, W_1 data were collected before the intervention began, and $W_2..W_9$ were collected during the intervention, and W_{10} data were collected one week post-intervention. We compute the change that occurs between week 1 and weeks 2 through 9.

ALGORITHM 1: Permutation-based Change Detection(X)

```

Input:  $X$  = time series data
Output: Vector  $V$  of (distance, significance) pairs
Initialize:  $n = 7$  # window length in number of days
Initialize:  $i = 1; j = i + n$ 
for each pair of windows,  $W_i, W_j \subseteq X$  do
     $W_i = X[i:i+n]$ 
     $W_j = X[j:j+n]$ 
     $k = 0; N = 1000$ 
     $C = \text{KL\_Divergence}(W_i, W_j)$ 
     $V^*$  = vector of length  $N$ 
    while  $k < N$  do
         $W_i^*, W_j^* = \text{Shuffle time intervals of } W_i \text{ and } W_j$ 
         $C^* = \text{KL\_Divergence}(W_i^*, W_j^*)$ 
         $V^* = \text{append}(V^*, C^*)$ 
         $k = k + 1$ 
    end while
     $sig = \text{BoxplotOutlierDetection}(C, V^*)$ 
     $V = \text{append}(V, (C, sig))$ 
     $j = j + n$ 
end for
return  $V$ 

```

We propose a small-window Permutation-based Change Detection approach to quantifying the amount of change, C , between two windows of time series sensor data. This algorithm is uniquely suited for analyzing activity-labeled sensor data. PCD is based on the notion of an activity curve [70], a compilation of $m=24$ probability distributions $R_t = \{d_{t,1}, d_{t,2}, \dots, d_{t,A}\}$

over A possible activities per time interval t in a day ($t = 1, 2, \dots, m$). Distribution $d_{t,i}$ represents the probability of activity A_i occurring during time interval t . Windows spanning multiple days (in our analysis, spanning one week) are averaged into an aggregate activity curve \hat{A} . To compute a distance between two activity curves \hat{A}_i and \hat{A}_j , the two curves are first aligned by dynamic time warping. Second, a symmetric version of KL divergence computes the distance between each pair of distributions at time interval t , as shown in Equation 3.

$$KL_{symmetric} = \sum_{k=1}^a d_{i,k} \cdot \log \frac{d_{i,k}}{d_{j,k}} + \sum_{k=1}^a d_{j,k} \cdot \log \frac{d_{j,k}}{d_{i,k}} \quad (3)$$

We adapt a test introduced by Ojala and Garriga [71] to determine whether the distance results are statistically significant. Statistical significance calculation is not easily available for prior methods such as uLSIF and RuLSIF. Our adaptation allows a permutation test to determine the significance of change scores calculated for a small number ($n=7$) of days. As shown in Algorithm 1, change score C is computed between windows W_i and W_j . Next, W_i and W_j are joined to form a new longer window. All time intervals within W are shuffled, then W is re-split into new windows, W_i^* and W_j^* . A change score C^* is computed between the new windows and added to a permutation-result vector V^* . The process is repeated N times. Finally, PCD compares C to the permutation vector V^* using boxplot outlier detection. Here, an outlier is defined as a value that is outside the interquartile range (75th percentile – 25th percentile) of V^* . The p-value estimated by computing the distance between C is an outlier, then the score is considered statistically significant. We employ boxplot outlier detection for this task because many alternative methods such as Grubb's test require that the data follow a normal distribution, an assumption which frequently does not apply to human behavior data. The output of the PCD algorithm is a vector V of change scores and associated significance vectors, each of which compares one week of observed human behavior data to the first baseline week.

6 EXPERIMENTAL RESULTS

We are interested in analyzing the impact of the B-Fit brain health intervention on the pattern of behaviors for the study participants. We start by analyzing one specific targeted behavior, then look at overall behavior. We hypothesize that study participants changed their behavior in response to the brain health intervention. We further postulate that these changes will be detected by PCD applied to activity-labeled smartwatch sensor data. Additionally, we want to determine whether overall behavior change in the intervention group will reflect a steadier increase over the nine weeks than for the non-intervention control group, consistent with the integration of new goals and maintenance of the earlier defined goals.

6.1 Validating PCD on Synthetic Data

First, we verify the reliability of our computational methods that apply PCD-based change detection to digital behavior markers as a method of quantifying change in overall behavior routine. To verify that PCD computes change scores as expected for digital marker data, we examine scores that are generated for synthetic data. We use real digital markers

from one randomly-selected intervention subject. We then add white Gaussian noise with monotonically-increasing values for the Gaussian distribution standard deviation. Table 4 verifies that the PCD-based change score increases monotonically with the amount of noise, or change, that is added to the real collected data.

6.2 Assessing Change in Target Activity

Next, we analyze the collected sensor data to determine how much change occurs for one of the targeted activities. Our primary goal for this paper is to determine whether our computational method can be used to detect and quantify change in overall behavior activity patterns as promoted by the intervention. However, we also posit that the same method can be used to detect change in a specific activity. The set of activities that are automatically recognized are listed in Section 2. The digital markers reflect these specific activities as well as parameters such as visited locations and movement statistics. Here, we pick a representative activity to demonstrate that the computational method can be used to analyze individual activities as needed.

We select *exercise* as the individual activity to analyze. One rationale for this choice is because the week discussing physical activity occurs early in the intervention. If behavior change is sustained per the intervention goal, change should be sustained throughout the data collection period in comparison with the week 0, pre-intervention baseline. A second motivation is that exercise represents a small set of possibilities for the participants (primarily walking and swimming). In contrast, the intervention topic that occurs the week before, cognitive engagement, is manifest through a large variety of individualized healthy behavior choices (e.g., play guitar, read a book, learn a new language), not all of which can be easily detected by smartwatch sensors. We include a subset of the digital markers for this analysis, namely, the amount of time spent per day on the automatically-labeled exercise activity, time of day that exercise occurs, the regularity of these activity on a daily basis, and statistical operations applied to these markers.

Table 5 summarizes the results of this analysis. The Permutation-based Change Detection values indicate that there is a greater change in Exercise habits for the intervention group than for the control group. These changes are statistically significant for more than half of the group. Because PCD provides change amounts without indicating the nature of the change, we also compute linear regression for exercise time and note that the slope for the intervention group is slightly higher than for the non-intervention group. In fact, the coefficient for the non-intervention group is close to zero, indicating that the PCD-detected changes may be due in some cases to natural variation in exercise regimens rather than intentional increase in exercise amount.

6.3 Assessing Overall Change in Behavior Routine

The analysis of change in exercise markers confirms that some behavior changes can be detected by wearable sensors and that our PCD algorithm can quantify these changes. These findings align with prior work that detects changes in exercise patterns [72]. On the other hand, we are particularly interested in observing the impact of a behavioral intervention on a person's entire pattern of behavior. Researchers have discovered interplay between

many behaviors, including but not limited to exercise and work [73], work, travel time, and social life [74], and diet and sleep [75]. Therefore, we hypothesize that engagement in the brain health intervention will impact overall behavior patterns. We analyze the complete set of digital markers collected for study participants and for the non-intervention group and compute PCD scores that compare the baseline week with the end-of-collection week.

The results of this analysis appear in Table 6. As anticipated, the changes are greater for the intervention group than for the control group. However, weekly changes were still fairly high for the non-intervention group. When we plot weekly change scores for individuals within the two groups, we gain new insights on the nature of the change. Almost all of the individuals in the intervention group exhibited consistently increasing change from baseline, as shown in Figure 3 (left). In contrast, the trends were highly varied for the non-intervention group. Figure 3 (right) provides one such example illustrating a common theme among non-intervention participants. To quantify these trends, we compute the linear regression coefficient for the change scores.

6.4 Assessing Change in Location

Next, we examine the widespread impact of a brain health intervention on a person's entire daily routine by assessing change in the types of locations where a person spends their time. To visualize the impact that intentional change can have on location patterns, we plotted the normalized, aggregated time spent in the home, outside the home (work, socializing, hobbies, or exercising) the week before intervention initiation and the last week of the intervention. Figure 4 shows this visualization for a 24-hour timeframe. As the graph shows, end-of-intervention routines include more time spent at home in the late-night and early-morning hours. One possible explanation for this is a change in behavior to incorporate more regular sleep hours and patterns. Furthermore, the midday hours reveal less time spent in travel (driving or riding other public transportation) and more time spent in other locations out of the home. This could potentially be attributed to incorporating more exercise, social times, and hobbies into the person's daily routine. Change in location markers between 0 and 9 were much greater for the intervention group, averaging 0.8469, while change for the non-intervention group averaged 0.3279.

6.5 Predicting Subject Group based on Delta Markers

In our final analysis, we employ a supervised learning algorithm to classify each subject into their corresponding group (intervention or control) based on digital behavior markers. Because we are focused on behavior changes related to the brain health intervention, we compute delta features to perform the classification. Specifically, we calculate the numeric change in each marker from week 0 to week 9. The set of delta features is mapped to the predicted subject group using a range of alternative supervised learning algorithms. To quantify the impact of the intervention on a person's overall activity routine, we compare the results using all delta features with those derived from using a subset of the available features. Specifically, we compute predictive accuracy using only delta features derived from movement-based sensors, labeled activity features, or location type features.

The results of this analysis are summarized in Table 7. We utilize a random forest with 100 trees and a varied number of features. Results are computed using leave-one-subject-out validation. As the table reflects, the delta values (calculated between the first and last weeks of data collection) were predictive of the subject group with an accuracy of 0.87 when all the features were included in the learned model. The ROC curve for this problem is shown in Figure 5. As the figure indicates, the corresponding AUC (area under the receive operating characteristic curve) is 0.95.

Using subsets of features still resulted in an accuracy that was better than random chance. However, considering changes in only movement features, activity features, or location features did not produce as strong results as considering changes in the entire set of features, corresponding the person's larger behavior routine context.

7 DISCUSSION AND CONCLUSION

In this study, we investigate methods to detect targeted and secondary behavior changes in response to a multi-modal brain health behavior intervention. By collecting continuous wearable sensor data, we can now monitor intervention impact in an objective, ecologically-valid manner. Through application of Permutation-Based Change Detection, we can quantify change in a subset of target behavior markers or in a person's overall pattern of behavior. For the brain health intervention, we did observe that targeted and secondary behaviors changed throughout the data collection process. While change was also observed for the non-intervention group, linear regression revealed that the amount of change steadily increased more for the intervention group than the non-intervention group. This may be an indication of intentional introduction of new, healthy behaviors into the subjects' daily routines.

The system we describe for detecting and quantifying behavior change relies on numerous interacting components. A limitation of this study is the error that can be introduced by each component in the system. The average self-reported compliance with the intervention was 1.7, falling between "partly met" and "completely met" participant-defined goals. As a result, not all intended healthy behaviors were regularly incorporated into the subjects' lives. Furthermore, weekly self-report may introduce error due to poor recall and objective assessment of behavior adherence. In future work, we will compare the reliability of smartwatch-detected behavior adherence, weekly paper-based self-reporting, and daily self-reporting using ecological momentary assessment techniques delivered through the watch interface.

While the smartwatches are designed to continuously collect data, there are gaps in the data collection when the participant did not charge and/or wear the watch. This may introduce error as a function of collecting data when a person is not wearing the watch or imputing missing data. The sensors themselves can occasionally report erroneous location or movement values [76], affecting the quality of the activity labels and digital markers. Similarly, the activity recognition model is based on a large sample of participant-labeled data. Because recognition accuracy was not perfect even on this sample, activity label error may be introduced into the final set of markers, as may error in labeling location categories.

The findings from this work have important health care implications. By observing behavior patterns before, during, and after a clinical intervention, we can better understand mechanisms and person-characteristics that lead to short-term and long-term change after an intervention. With continued development, this introduced architecture will provide a more continuous, objective measurement of the impact of behavioral interventions on changes in behavior patterns, something that has been difficult to capture clinically but is an important part of interventions. These insights can help design more adaptive, efficacious, community-based preventive interventions for midlife and older adults at risk of developing dementia.

ACKNOWLEDGMENTS

This work is supported in part by an H.L. Eastlick Distinguished Professorship; National Institutes of Health grants R01EB009675 and R25AG046114 and by National Science Foundation grant 1954372.

REFERENCES

- [1]. U.S. Department of Health and Human Services, “Healthy People 2020,” Washington, D.C., 2015.
- [2]. Aung MH, Matthews M, and Choudhury T, “Sensing behavioral symptoms of mental health and delivering personalized interventions using mobile technologies,” *Depression and Anxiety*, vol. 34, no. 7, pp. 603–609, 2017. [PubMed: 28661072]
- [3]. Mishra V, “From sensing to intervention for mental and behavioral health,” in *UbiComp*, 2019, pp. 388–392.
- [4]. Das B, Cook DJ, Krishnan N, and Schmitter-Edgecombe M, “One-class classification-based real-time activity error detection in smart homes,” *IEEE Journal of Selected Topics in Signal Processing*, 2016.
- [5]. Boyd B, McAlister C, Arrotta K, and Schmitter-Edgecombe M, “Self-reported behavior change and predictors of engagement with a multidomain brain health intervention for midlife and older adults: A pilot clinical trial,” under review, 2021.
- [6]. Chen K, Zhang D, Yao L, Guo B, Yu Z, and Liu Y, “Deep learning for sensor-based human activity recognition: Overview, challenges and opportunities,” *Journal of the ACM*, vol. 37, no. 4, p. 111, 2020.
- [7]. Bulling A, Blanke U, and Schiele B, “A tutorial on human activity recognition using body-worn inertial sensors,” *ACM Computing Surveys*, vol. 46, no. 3, pp. 107–140, 2014.
- [8]. Asim Y, Azam MA, Ehatisham-ul-Haq M, Naeem U, and Khalid A, “Context-aware human activity recognition (CAHAR) in-the-wild using smartphone accelerometer,” *IEEE Sensors*, vol. 8, pp. 4361–4371, 2020.
- [9]. Tian Y, Zhang J, Chen L, Geng Y, and Wang X, “Selective ensemble based on extreme learning machine for sensor-based human activity recognition,” *Sensors*, vol. 19, no. 16, p. 3468, 2019.
- [10]. Nazabal A, Garcia-Moreno P, Artes-Rodriguez A, and Ghahramani Z, “Human activity recognition by combining a small number of classifiers,” *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1342–1351, 2016. [PubMed: 26208368]
- [11]. Wang J, Chen Y, Hao S, Peng X, and Hu L, “Deep learning for sensor-based activity recognition: A survey,” *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.
- [12]. Hammerla NY, Halloran S, and Ploetz T, “Deep, convolutional, and recurrent models for human activity recognition using wearables,” in *International Joint Conference on Artificial Intelligence*, 2016.
- [13]. Ploetz T and Guan Y, “Deep learning for human activity recognition in mobile computing,” *Computer*, vol. 51, no. 5, pp. 50–59, 2018.
- [14]. Guan Y and Ploetz T, “Ensembles of deep LSTM learners for activity recognition using wearables,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, p. 11, 2017.

- [15]. Marshall GA, Rentz DM, Frey MT, Locascio JJ, Johnson KA, and Sperling RA, "Executive function and instrumental activities of daily living in mild cognitive impairment and Alzheimer's disease," *Alzheimer's and Dementia*, vol. 7, pp. 300–308, 2011.
- [16]. Miller LS, Brown CL, Mitchell MB, and Williamson GM, "Activities of daily living are associated with older adult cognitive status: Caregiver versus self-reports," *Journal of Applied Gerontology*, vol. 32, pp. 3–30, 2013. [PubMed: 25473924]
- [17]. Bharti P, De D, Chellappan S, and Das SK, "HuMAN: Complex activity recognition with multi-modal multi-positional body sensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 4, pp. 857–870, 2019.
- [18]. Peng L, Chen L, Wu M, and Chen G, "Complex activity recognition using acceleration, vital sign, and location data," *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1488–1498, 2019.
- [19]. Lin M and Hsu W-J, "Mining GPS data for mobility patterns: A survey," *Pervasive and Mobile Computing*, vol. 12, pp. 1–16, 2014.
- [20]. Boukhechba M, Huang Y, Chow P, Fua K, Teachman BA, and Barnes LE, "Monitoring social anxiety from mobility and communication patterns," *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 749–753, 2017.
- [21]. Kankanhalli A, Saxena M, and Wadhwa B, "Combined interventions for physical activity, sleep, and diet using smartphone apps: A scoping literature review," *International Journal of Medical Informatics*, vol. 123, pp. 54–67, 2019. [PubMed: 30654904]
- [22]. Sprint G and Cook DJ, "Unsupervised detection and analysis of changes in everyday physical activity data," *Journal of Biomedical Informatics*, 2016.
- [23]. de Zambotti M, Cellini N, Goldstone A, Colrain IM, and Baker FC, "Wearable sleep technology in clinical and research settings," *Medicine and Science in Sports and Exercise*, vol. 51, no. 7, pp. 1538–1557, 2019. [PubMed: 30789439]
- [24]. Wang R et al., "Tracking depression dynamics in college students using mobile phone and wearable sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–26, 2018.
- [25]. Jain A, Popescu M, Keller J, Rantz M, and Markway B, "Linguistic summarization of in-home sensor data," *Journal of Biomedical Informatics*, vol. 96, p. 103240, 2019. [PubMed: 31260752]
- [26]. Wang W et al., "Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, p. 141, 2018.
- [27]. Costa J, Guimbretier F, Jung MF, and Choudhury T, "BoostMeUp: Improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019, p. 40.
- [28]. Stanley MA et al. , "The peaceful mind program: A pilot test of a cognitive-behavioral therapy-based intervention for anxious patients with dementia," *The American Journal of Geriatric Psychiatry*, vol. 21, no. 7, pp. 696–708, 2013. [PubMed: 23567399]
- [29]. Carreiro S, Newcomb M, Leach R, Ostrowski S, Boudreaux ED, and Amante D, "Current reporting of usability and impact of mHealth interventions for substance use disorder: A systematic review," *Drug and Alcohol Dependence*, vol. 215, p. 108201, 2020. [PubMed: 32777691]
- [30]. Liu J et al. , "Bayesian structural time series for biomedical sensor data: A flexible modeling framework for evaluating interventions," *bioRxiv*, 2020.
- [31]. Dhana K, Evans DA, Rajan KB, Bennett DA, and Morris MC, "Healthy lifestyle and the risk of Alzheimer dementia: Findings from 2 longitudinal studies," *Neurology*, vol. 95, no. 4, pp. e374–e383, 2020. [PubMed: 32554763]
- [32]. Alberdi Aramendi A et al. , "Smart home-based prediction of multi-domain symptoms related to Alzheimer's Disease," *IEEE Journal of Biomedical and Health Informatics*, 2018, doi: 10.1109/JBHI.2018.2798062.

- [33]. Li J, Rong Y, Meng H, Lu Z, Kwok T, and Cheng H, “TATC: Predicting Alzheimer’s disease with actigraphy data,” in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018, pp. 509–518.
- [34]. Aminikhanghahi S and Cook DJ, “A survey of methods for time series change point detection,” Knowledge and Information Systems, vol. 51, no. 2, pp. 339–367, Sep. 2017, doi: 10.1007/s10115-016-0987-z. [PubMed: 28603327]
- [35]. Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, and Muller P-A, “Deep learning for time series classification: A review,” Data Mining and Knowledge Discovery, vol. 33, no. 4, pp. 917–963, 2019.
- [36]. Cho H and Fryzlewicz P, “Multiple-change-point detection for high dimensional time series via sparsified binary segmentation,” Journal of the Royal Statistical Society, vol. 77, no. 2, pp. 475–507, 2015.
- [37]. Yamanishi K and Takeuchi J, “A unifying framework for detecting outliers and change points from non-stationary time series data,” in ACM SIGKDD international Conference on Knowledge Discovery and Data Mining, Jul. 2002, p. 676, doi: 10.1145/775047.775148.
- [38]. Liu S, Yamada M, Collier N, and Sugiyama M, “Change-point detection in time-series data by relative density-ratio estimation.,” Neural networks : the official journal of the International Neural Network Society, vol. 43, pp. 72–83, Jul. 2013, doi: 10.1016/j.neunet.2013.01.012. [PubMed: 23500502]
- [39]. Livingston G et al. , “Dementia prevention, intervention, and care: 2020 report of the Lancet Commission,” The Lancet, vol. 396, p. 10248, 2020.
- [40]. Andrieu S et al. , “Effect of long-term omega 3 polyunsaturated fatty acid supplementation with or without multidomain intervention on cognitive function in elderly adults with memory complaints (MAPT): A randomized, placebo-controlled trial,” The Lancet Neurology, vol. 16, no. 5, pp. 377–389, 2017. [PubMed: 28359749]
- [41]. Rosenberg A et al. , “Multidomain lifestyle intervention benefits a large elderly population at risk for cognitive decline and dementia regardless of baseline characteristics: The FINGER trial,” Alzheimer’s & Dementia, vol. 14, no. 3, pp. 263–270, 2018.
- [42]. Hartmann C, Dohle S, and Siegrist M, “A self-determination theory approach to adults’ healthy body weight motivation: A longitudinal study focusing on food choices and recreational physical activity,” Psychology and Health, vol. 30, no. 8, pp. 924–948, 2015. [PubMed: 25584714]
- [43]. Huang CL, Yang SC, and Chiang CH, “The associations between individual factors, eHealth literacy, and health behaviors among college students,” Journal of Medical Internet Research, vol. 19, no. 1, p. e15, 2017. [PubMed: 28073739]
- [44]. Norton S, Matthews F, Barnes DE, Yaffe K, and Brayne C, “Potential for primary prevention of Alzheimer’s disease: an analysis of population-based data,” Lancet Neurology, vol. 13, no. 8, pp. 788–794, 2014. [PubMed: 25030513]
- [45]. National Academies of Sciences, Engineering and Medicine, “Preventing cognitive decline and dementia: A way forward,” 2017.
- [46]. Phillips C, “Lifestyle modulators of neuroplasticity: How physical activity, mental engagement, and diet promote cognitive health during aging,” Neural Plasticity, p. 3589271, 2017. [PubMed: 28695017]
- [47]. Samadi M, Moradi S, Moradinazar M, Mostafai R, and Pasdar Y, “Dietary pattern in relation to the risk of Alzheimer’s disease: A systematic review,” Neurological Sciences, vol. 40, no. 2031–2043, 2019. [PubMed: 31240575]
- [48]. Venkatraman V, Sanderson A, Cox KL, Ellis K, Steward C, and Phal PM, “Effect of a 24-month physical activity program on brain changes in older adults at risk of Alzheimer’s disease: The AIBL active trial,” Neurobiology of Aging, vol. 89, pp. 132–141, 2020. [PubMed: 31324405]
- [49]. Kilmova B, Maresova P, and Kamil K, “Alzheimer’s Disease: Physical activities as an effective intervention tool - A mini-review,” Current Alzheimer Research, vol. 16, no. 2, pp. 166–171, 2019. [PubMed: 30277151]
- [50]. Chen J, Duan Y, Li H, Lu L, Liu J, and Tang C, “Different durations of cognitive stimulation therapy for Alzheimer’s disease: A systematic review and meta-analysis,” Clinical Interventions in Aging, vol. 14, pp. 1243–1254, 2019. [PubMed: 31371930]

- [51]. Cherry KE, Jackson Walker EM, Silva Brown JL, Volaufova J, LaMotte LR, and Su LJ, “Social engagement and health in younger, older, and oldest-old adults in the Louisiana Healthy Aging Study (LHAS),” *Journal of Applied Gerontology*, vol. 32, pp. 51–75, 2013. [PubMed: 23526628]
- [52]. Lin J-FL, Imada T, and Kuhl PK, “Neuroplasticity, bilingualism, and mental mathematics: A behavior-MEG study,” *Brain and Cognition*, vol. 134, pp. 122–134, 2019. [PubMed: 30975509]
- [53]. Wahl D et al. , “Aging, lifestyle and dementia,” *Neurobiology of Disease*, vol. 130, p. 104481, 2019. [PubMed: 31136814]
- [54]. Schmitter-Edgecombe M and Dyck D, “A cognitive rehabilitation multi-family group intervention for individuals with mild cognitive impairment and their care-partners,” *Journal of the International Neuropsychological Society*, 2014.
- [55]. Park CS, Troutman-Jordan M, and Nies MA, “Brain health knowledge in community-dwelling older adults,” *Educational Gerontology*, vol. 38, no. 9, pp. 650–657, 2012.
- [56]. Guerrero APS et al. , “Rural mental health training: An emerging imperative to address health disparities,” *Academic Psychiatry*, vol. 43, pp. 1–5, 2019. [PubMed: 30535843]
- [57]. Brandt J and Folstein M, *Telephone Interview for Cognitive Status*. Lutz, FL: Psychological Assessment Resources, Inc., 2003.
- [58]. Wilson G, Doppa J, and Cook DJ, “Multi-source deep domain adaptation with weak supervision for time-series sensor data,” in *KDD*, 2020.
- [59]. Wang J, Zheng VW, Chen Y, and Huang M, “Deep transfer learning for cross-domain activity recognition,” in *International Conference on Crowd Science and Engineering*, 2018, p. 16.
- [60]. Rokni SA and Ghasemzadeh H, “Autonomous training of activity recognition algorithms in mobile sensors: A transfer learning approach in context-invariant views,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 8, 2018.
- [61]. Ma Y and Ghasemzadeh H, “LabelForest: Non-parametric semi-supervised learning for activity recognition,” in *AAAI Conference on Artificial Intelligence*, 2019.
- [62]. Miu T, Missier P, and Plotz T, “Bootstrapping personalised human activity recognition models using online active learning,” in *IEEE International Conference on Computer and Information Technology*, 2015.
- [63]. Banos O, Galvez J-M, Damas M, Pomares H, and Rojas I, “Window size impact in human activity recognition,” *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014. [PubMed: 24721766]
- [64]. Haresamudram H, V Anderson D, and Ploetz T, “On the role of features in human activity recognition,” in *International Symposium on Wearable Computers*, 2019, pp. 78–88.
- [65]. Nosakhare E and Picard R, “Toward assessing and recommending combinations of behaviors for improving health and well-being,” *ACM Transactions on Computing for Healthcare*, vol. 1, no. 1, p. 4, 2020.
- [66]. Obuchi M et al. , “Predicting brain functional connectivity using mobile sensing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, p. 23, 2020.
- [67]. Cook DJ, Schmitter-Edgecombe M, Jonsson L, and V Morant A, “Technology-enabled assessment of functional health,” *IEEE Reviews in Biomedical Engineering*, p. to appear, 2018.
- [68]. Li D et al. , “Feasibility study of monitoring deterioration of outpatients using multimodal data collected by wearables,” *ACM Transactions on Computing for Healthcare*, vol. 1, no. 1, p. 5, 2020.
- [69]. Depner CM et al. , “Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions,” *Sleep*, vol. 43, no. 2, p. zsz254, 2020. [PubMed: 31641776]
- [70]. Dawadi PN, Cook DJ, and Schmitter-Edgecombe M, “Modeling patterns of activities using activity curves,” *Pervasive and Mobile Computing*, vol. 28, 2016, doi: 10.1016/j.pmcj.2015.09.007.
- [71]. Ojala M and Garriga CG, “Permutation tests for studying classifier performance,” *Journal of Machine Learning Research*, vol. 11, pp. 1833–1863, 2010.
- [72]. Hu B and Charniak T, “Wearable technological platform for multidomain diagnostic and exercise interventions in Parkinson’s disease,” *International Review of Neurobiology*, vol. 147, pp. 75–93, 2019. [PubMed: 31607363]

- [73]. Jindo T, Kai Y, Kitano N, Tsunoda K, Nagamatsu T, and Arao T, "Relationship of workplace exercise with work engagement and psychological distress in employees: A cross-sectional study from the MYLS study," *Preventive Medicine Reports*, vol. 17, p. 101030, 2020. [PubMed: 31890476]
- [74]. Nisic N and Kley S, "Gender-specific effects of commuting and relocation on a couple's social life," *Demographic Research*, vol. 40, no. 36, pp. 1047–1062, 2020.
- [75]. Binks H, Vincent GE, Gupta C, Irwin C, and Khalesi S, "Effects of diet on sleep: A narrative review," *Nutrients*, vol. 12, no. 4, p. 936, 2020.
- [76]. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook DJ, and Ghasemzadeh H, "How accurate is your activity tracker? A comparative study of step counts in low-intensity physical activities," *Journal of Medical Internet Research*, 2017.

CCS CONCEPTS

- Ubiquitous and mobile computing systems and tools
- Machine learning approaches
- Health informatics

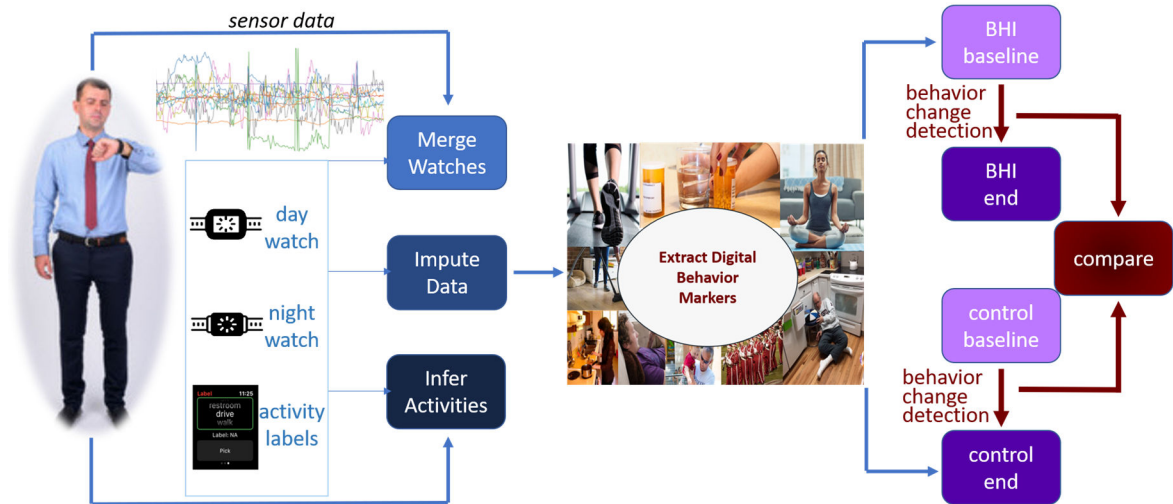


Figure 1: Process of detecting behavior change with smartwatches. Data are continuously collected, merged, completed, and labeled with inferred activities. From these data, digital markers are extracted for each hour and day. We quantify the amount of change from begin to end of data collection for the experiment and non-intervention groups using behavior change detection and compare amount of change between the two groups.

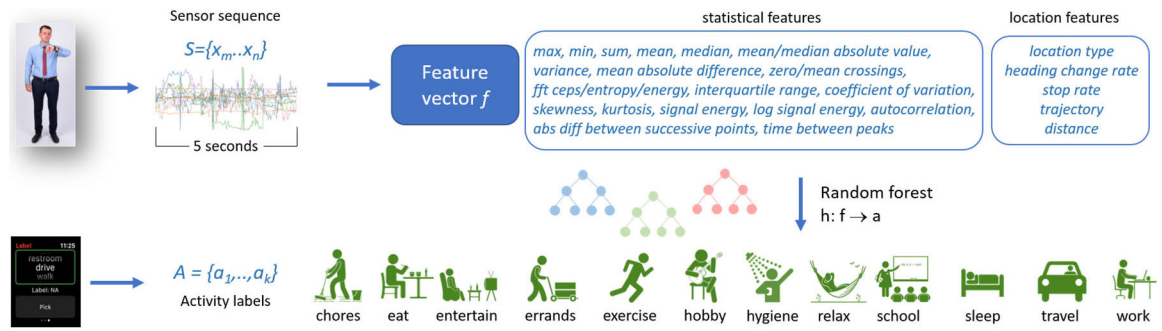


Figure 2: Activity labeling. A feature vector is extracted from five seconds of sensor data. A random forest maps the feature vector onto one of the activity choices, based on training data provided in real time by smartwatch users.

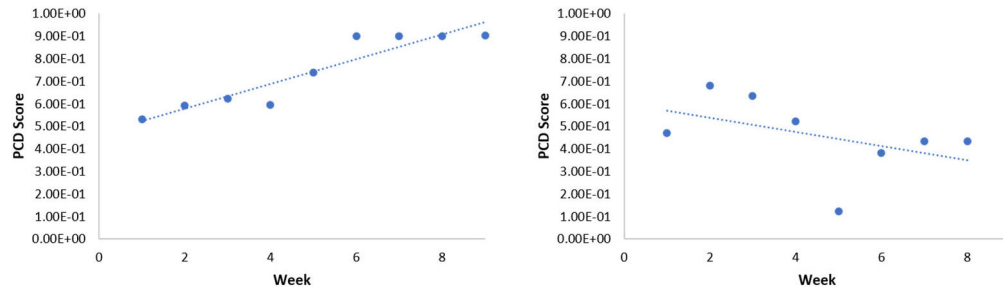


Figure 3: PCD score trends. (left) A plot of the PCD scores for a subject in the intervention group. (right) A plot of PCD scores for a subject in the non-intervention group.

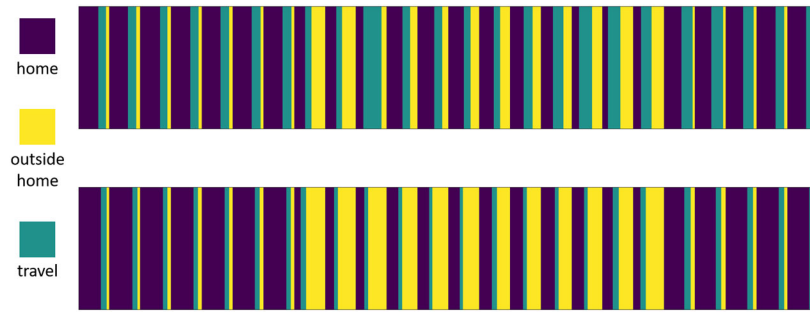


Figure 4: Visualization of time spent at location types for the intervention group. The x axis shows a 24-hour time period starting at midnight. Bar sizes indicate the aggregated, normalized duration spent at the corresponding location type.

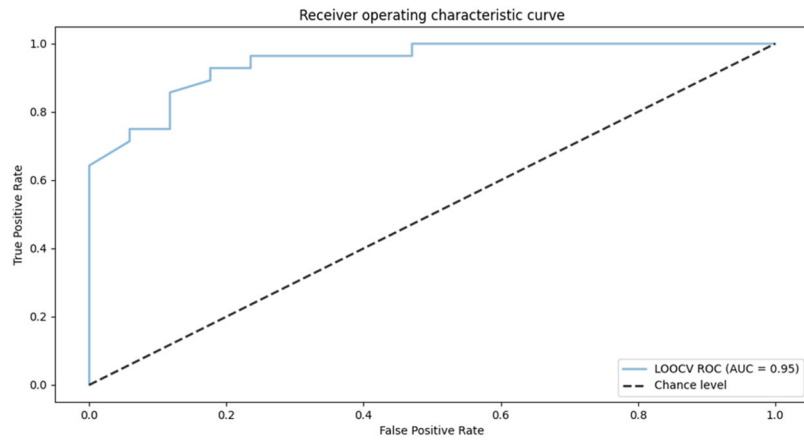


Figure 5: ROC curve for random forest-based prediction of subject group, using pre- and post-intervention changes in digital markers as predictive features. The corresponding AUC is 0.95.

Table 1:

Didactic topics discussed in group sessions

Session	Education Materials Discussed
Session 1	Overview of the brain, cognitive aging, MCI, and dementia
Session 2	Cognitive engagement
Session 3	Physical activity and cardiovascular risk factors
Session 4	Nutrition and diet
Session 5	Social engagement
Session 6	Sleep and stress reduction
Session 7	Compensatory strategies and assistive technology

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Participant demographics

Group	Age (Mean / SD)	Years of Education (Mean / SD)	Male	Female	Caucasian	Other Ethnicity
<i>BHI (n=28)</i>	63.18 / 9.19	16.00 / 1.86	5	25	26	4
<i>NI (n=17)</i>	67.22 / 15.89	15.89 / 2.49	4	13	16	1

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Behavioral markers

Type	Features
Daily	Amount of time (in minutes) spent on each activity Time of day (hours and minutes past midnight) of first and last occurrence for each activity Amount of time spent at each location type Step count, activity level, distance traveled Number of missing values (minutes with no sensor readings)
Hourly	Amount of time spent on each activity type and each location type Total rotation, acceleration, distance traveled
Overall	Regularity index (within week, within weekdays, between weeks) Diurnal and Circadian rhythm

Table 4:

PCD values for digital markers with white additive Gaussian noise.

σ			
0.0	0.1	0.2	0.3
0.0005	0.8243	0.8510	0.8908

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5:

Change in Exercise activity from baseline to end of intervention. The table summarizes the average PCD value for the intervention group and the control group together with the number of persons that experienced a statistically significant change. The table also summarizes the average linear regression coefficient and R value for each of the groups.

PCD-intervention (%sig)	PCD-control (%sig)	Coefficient-intervention (R)	Coefficient-control (R)
0.3500 (53.57%)	0.1478 (29.41%)	0.1016 (0.25)	0.0008 (0.40)

Table 6:

Change in overall activity routine from baseline to end of intervention. The table summarizes the average PCD value for the intervention group and the control group together with the number of persons that experienced a statistically significant change. The table also summarizes the average linear regression coefficient and R value for each of the groups.

PCD-intervention (%sig)	PCD-control (%sig)	Coefficient-intervention (R)	Coefficient-control (R)
0.7460 (50.00%)	0.4428 (17.64%)	0.1598 (0.76)	0.0230 (0.37)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7:

Accuracy of predicting whether a subject is in the intervention group or control group using a random forest. Results for are reported for leave-one-subject-out classification and for alternative sets of features.

movement	activity	location	all
0.76	0.73	0.80	0.87

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript