

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Fusing Ambient and Mobile Sensor Features into a Behaviorome for Predicting Clinical Health Scores

Diane J. Cook¹, Fellow, IEEE, and Maureen Schmitter-Edgecombe²

¹School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164 USA

²Department of Psychology, Washington State University, Pullman, WA 99164 USA

Corresponding author: Diane J. Cook (e-mail: djcook@wsu.edu).

This work was supported in part by the National Institutes of Health under Grants R01AG065218, R01NR016732, R01EB009675, and by the National Science Foundation under Grant 1954372.

ABSTRACT Advances in machine learning and low-cost, ubiquitous sensors offer a practical method for understanding the predictive relationship between behavior and health. In this study, we analyze this relationship by building a behaviorome, or set of digital behavior markers, from a fusion of data collected from ambient and wearable sensors. We then use the behaviorome to predict clinical scores for a sample of $n=21$ participants based on continuous data collected from smart homes and smartwatches and automatically labeled with corresponding activity and location types. To further investigate the relationship between domains, including participant demographics, self-report and external observation-based health scores, and behavior markers, we propose a joint inference technique that improves predictive performance for these types of high-dimensional spaces. For our participant sample, we observe correlations ranging from small to large for the clinical scores. We also observe an improvement in predictive performance when multiple sensor modalities are used and when joint inference is employed.

INDEX TERMS activity recognition; behavior markers; joint inference; smart homes; smartwatches

I. INTRODUCTION

Care providers, patients, families, and researchers all struggle to understand human health and its influences. Until recently, theories that connect behavior and health were based on self-report, which can suffer from retrospective memory limitations [1], or experimenter observations, which can introduce confounds and unintended bias [2]. Because we can now collect continuous behavior and health data in an ecologically-valid manner, we can leverage substantial evidence to support or contradict these theories. Despite the tremendous potential of emerging technologies to monitor and assess health and behavior, many of them are constrained to operate in controlled environments, laboratory settings, and clinics.

We hypothesize that multi-modal sensor data collected in uncontrolled settings can be harnessed to predictively connect health and behavior. 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, we will investigate whether clinical scores can be predicted from time series data. If successful, such

predictions can provide unprecedented, naturalistic insight into a person's health status and the relationship between behavior and health. They will allow caregivers to track and predict health trajectories over time, which may be particularly valuable for individuals going through rehabilitation or experiencing cognitive decline. Sensor-based continual assessment of daily physical and behavior patterns, combined with predicted clinical scores, provides insights that cannot be gleaned from spending a few minutes in a doctor's office or clinic.

The specific research contributions we make in this paper are the following. First, we describe methods of labeling ambient sensor and wearable sensor data in real time with corresponding activity labels. Second, we introduce methods to extract a behaviorome, or set of digital behavior markers, from activity-labeled time series data. Third, we investigate the predictability of clinical scores from a person's behaviorome. In this investigation, we collect continuous smart home and smartwatch data for $n=21$ healthy older adults while they perform their normal daily routines. Finally, we

introduce a method to enhance clinical score predictability by leveraging the relationships between multiple target variables.

II. RELATED WORK

Sensor data hold the potential to give researchers substantial information about a person's behavior as well as health. While sensors have become low cost, wireless, and deployable in real-world settings, researchers have found no single "silver bullet" sensor that provides all the necessary insight into a person's health and behavior. However, researchers have made progress in linking features of individual sensor modalities with specific health characteristics. Here we review recent advances in analyzing time series sensor data to gain insights for automated health assessment.

Our study combines behavior markers from wearable sensors and ambient sensors. Other sensors could be considered for this task as well, such as video, audio, phone, and biometric, to name a few. We restrict our analysis to sensors which are commercially available, easily deployable, and able to handle a variety of environment conditions. Within these constraints, researchers have made considerable progress in performing health assessment from sensor data. One such sensor platform is provided by wearable and mobile sensors. Wearable sensors typically quantify aspects of movement. These sensors may be embedded into inertial measurement units (IMUs) or included on commercial phones and watches which collect acceleration and angular velocity data.

In one study, Mancini et al. [3] analyzed wearable data to analyze movement parameters when a person performed turning motions. Customized wearable devices collected angular velocities. These velocities were shown to be predictive of falls and of cognitive function as determined by Clinical Dementia Rating scores. In other work, Adler et al. [4] collected data from a commercial smartphone to predict schizophrenia-related psychotic relapse using a neural network. They extracted features from movement-based sensors but also included app use, phone calls, and inferred sleep times. In addition to these features, Wang et al. [5] also inferred movement category (e.g., walking, in a vehicle) and computed an index that estimates the regularity of a person's daily schedule from phone-based sensors. They used this information to infer symptoms of depression as indicated by in-the-moment Patient Health Questionnaire-4 (PHQ-4). Huckins et al. [6] utilized a similar set of features to predict PHQ-4-detected anxiety among college students at the beginning of the COVID-19 pandemic. Focusing on a different health domain, Newland et al. [7] analyzed movement data collected from wearable sensors to infer clinical scores that are specific for assessing symptoms of multiple sclerosis.

In other studies, researchers utilized information from wearable sensors to infer a behavior or health marker that is consistent with a larger health condition. For example, Fiorini et al. [8] analyzed statistical features from wearable

accelerometer and angular velocity readings to measure a person's auditory sustained attention. This measure, in turn, provided insights on the person's cognitive function and thus the impact of cognitive training on cognitive function. Fallahzadeh et al. [9] collected movement and location information on smartphones to predict and promote medication adherence. They extracted statistical features from movement-based sensors, labeled data with corresponding activities categories, and used this information to infer whether a patient would be adherent to their medication schedule at the prescribed time based on their activity context.

Wearable sensors are particularly insightful at analyzing and inferring movement-based parameters that help to determine a person's health status. In particular, Ravichandran et al. [10] surveyed and compared the effectiveness of alternative commercial accelerometer-based devices for detecting sleep stages and durations. Fallmann and Chen [11] also surveyed approaches to detect and analyze sleep stages from wearable sensor data, focusing on research advances. Lee et al. [12] analyzed smartphone motion sensors to measure walking movements of healthy individuals and an individual with back pain. In particular, they calculated the variation of rotation about three axes and found that the participant with back pain demonstrated a wider range of variations, indicating a diminished level of walking balance.

The above studies primarily analyze movement sensors to provide insights on health status and behavior parameters that impact health status. However, some recent work provides evidence that location information can also be gathered from mobile devices and generalized among individuals. Lin and Hsu [13] introduced multiple generalizable location features, including heading change rate, distance covered, and velocity change rate. They also suggested clustering location readings to detect person-specific frequented spots. Boukhechba et al. [14] employed this strategy to identify significant locations. They used this information together with phone call frequency to predict a person's social anxiety.

Similar to these reviewed studies, we extract features from both movement and location information provided by wearable devices. Wearable devices offer other data modalities as well. These include phone usage, heart rate, and even audio and video data. These data sources offer valuable behavior and health features. For example, Tseng et al. [15] found that they could predict a person's inhibitory control based solely on phone usage statistics. To enable longitudinal data collection without additional hardware and participant involvement, we do not incorporate these additional wearable information sources into the analysis.

Ambient sensors that are embedded into physical environments complement wearable sensors by localizing a person within the building. This information, combined with environment state readings, can be used to longitudinally track complex activity patterns. As in the case of analyzing wearable sensors, researchers have devised methods to extract features from longitudinal ambient sensor data that

provide insights on behavior patterns. For example, Austin et al. [16] inferred behavior parameters from passive infrared motion sensors. The parameters include time spent out of the home, indoor walking speed, and time spent sleeping. From these features, they were able to predict loneliness among older adults. Akl et al. [17] also analyzed ambient motion sensor data to calculate times that individuals spent in rooms of the house. Based on this information, they were able to infer a person’s cognitive health as indicated by Mini-Mental State Examination and Clinical Dementia Rating scores. Dawadi et al. [18] similarly extracted features from motion and door sensors. From these features, they inferred activity categories and mapped activity features (duration and time of day) onto clinical cognitive health indicators including the Repeatable Battery of Neuropsychological Status and Timed Up and Go scores. Aramendi et al. [19] collected similar activity features over multiple years to predict these scores as well as scores on a self-report measure of instrumental activities of daily living (IADL-C).

Fritz et al. [20] collected similar features from ambient sensors as the previous studies. In this case, however, the features were used to detect pain episodes for participants with chronic health conditions. Skubic et al. [21] also collected data in naturalistic settings to identify possible health events. Unlike the previous studies, they analyzed video data together with ambient sensors, and created rules to identify anomalies from extracted features that increased the possibility of detecting events requiring intervention. In our study, we include features collected from motion, door, light, and temperature ambient sensors. As indicated by these prior studies, such sensors can be used to recognize activities and in turn provide insights on health status.

In all of these prior works, as in this current study, the goal of the technology is monitoring behavior over time, in the wild, to perform health assessment. In contrast to prior work, this study extracts a large set of behavior markers from both wearable and ambient sensor data. Because sensor platforms provide distinct insights in these routines, we hypothesize that the fused information will provide a more comprehensive picture of behavior patterns and thus improve predictive performance over a range of clinical health scores. Additionally, we introduce a method to predict multiple health scores in a way that boosts the prediction of each health parameter. These insights can be used to better understand the impact of behavior on health and to inform the continued design of assessment measures.

II. METHODS

A. DATA COLLECTION

In this paper, we analyze smart home data collected from 21 healthy older adult volunteers (16 female, 5 male). The mean age was 69 (SD 10.9) and the distribution of demographics is shown in Fig.1. All participants were recruited through community advertisements or referral from physicians and

local community agencies. Before they were selected for the study, participants completed a health screening interview and the Telephone Interview of Cognitive Status [22] over the phone. The interview allowed experimenters to rule out exclusion criteria. In particular, we enrolled participants who were at least 45 years of age and did not indicate a possibility of dementia or lack of insight that could lead to unreliable self-report data (TICS ≥ 26). The data were collected as part of a larger study focused on developing and validating methods for health assessment and intervention that utilize continuously-collected sensor data in naturalistic, real-world settings. The parent study does not target specific distributions of education levels, age, gender, or race.

We equipped the homes of these study participants with a CASAS “smart home in a box” [23]. Each CASAS smart home apartment contains ambient sensors, consisting of passive infrared motion/light sensors attached to the ceilings and door/temperature sensors attached to external doors and commonly-used cabinets. In participant homes, experimenters install approximately two sensors per room with adhesive strips that can be removed at the end of data collection. Data can be collected from these sensors for long periods of time while requiring no extra tasks from the participants, and the efficacy of these sensors to capture activity patterns and indicate health status has been demonstrated in prior work [19], [24], [25]. The CASAS sensors generate timestamped readings whenever the sensed state changes (e.g., still to motion, door closed to open). We collected 1 month of ambient sensor data for each participant. A 1-month data collection period was chosen to ensure that routine behavior patterns that occur on a daily, weekly, or monthly basis would be observed. This study was approved by the Washington State University Institutional Review Board.

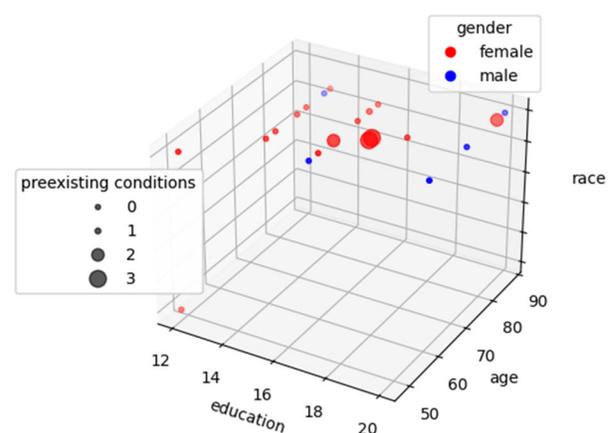


FIGURE 1. Distribution of study participants based on years of education, race, age, gender, and the number of preexisting conditions in these categories: neurological disability, head injury, and experiencing changes in memory performance.

During the same month, we also collected continuous smartwatch data for each participant. Participants wore the watch continuously while data were collected at 1Hz. To provide continuous data collection, participants wore one watch during the day while charging a second watch, then wore the other watch at night while charging the first. This sample rate was selected to ensure that data could be collected continuously during the day without draining the battery. Consistent with prior studies [3]–[8], smartwatch data consist of timestamped readings for 3D accelerometer, 3D gyroscope, course, and speed. We additionally collected location data. As suggested in earlier work [13], [14], raw location coordinates were not included in the models because of the privacy risk and because they do not generalize well to new participants. Instead, we employ an open street map to map each coordinate onto a location type (home, road, work, other).

Prior activity recognition and health assessment research offers evidence that movement-based sensors including accelerometers and gyroscopes provide insight on behavior patterns “in the wild” [26]–[28]. Many of these previous efforts analyze movement-based activities such as walking, lying down, and exercising. Recognizing more complex activities such as working and shopping relies on additional information, because the movements associated with sitting in front of a television at home will be similar to the movements associated with sitting in a meeting at work. We incorporate location information to aid in detecting and distinguishing these more complex activities.

We also calculated person-specific location parameters including mean location and the top three most-frequented location regions by time of day (morning, afternoon, evening, night). For some participants, there were periods of time during transition between watches where smartwatch data were missing (mean minutes of missing readings/day=5.59, s.d.=5.33). We imputed the missing values using random forest regression with 100 trees. The features employed for imputation-based prediction are the previous readings from all sensors leading up to the missing time and readings for the missing sensor occurring at the same time of day (number of minutes past midnight) for the seven prior days for which readings are available. In the case of ambient sensor data from smart homes, no data were missing. These data collection environments require no actions on the part of participants. Because batteries for each sensor last for more than a year without a need for recharge, data collection did not experience any interruptions in the smart homes.

Finally, a neuropsychologist team collected self-report information for each participant and administered a battery of clinical assessments. We include a total of 129 self-report answers to health-related items and clinical parameters for analysis. Baseline self-report questions primarily assessed everyday health-related behaviors (e.g., level of engagement in physical exercise, educational community activities) based on a Likert Scale. We also collected self-report answers and performance on a cognitive n-back test [29] during data

TABLE I
TARGETED CLINICAL SCORES.

<i>Measure</i>	<i>Construct Assessed</i>
Wechsler Test of Adult Reading (WTAR)	Premorbid verbal intellectual abilities
Telephone Interview of Cognitive Status (TICS)	Global cognitive status
Repeatable Battery of Neuropsychological Status (RBANS)	General neurocognitive status
F-A-S Test (FAS)	Executive function
Timed Up and Go Test (TUG)	Functional mobility
Dysexecutive questionnaire (DEX)	Self-reported everyday executive dysfunction
Alzheimer’s disease Cooperative Study Activities of Daily Living Inventory (ADCS-ADL)	Self-reported everyday activity ability and functional independence

collection using an ecological momentary assessment (EMA) app. EMA self-report questions assessed in-the-moment functioning, socialization, physical and mental activity, fatigue, and mood. For the purposes of this work, these scores were averaged across the data collection period. The neurocognitive measures assessed performance across varying cognitive domains (e.g., memory, executive function, attention).

We selected 7 clinical scores as target variables for the prediction models. The target scores are summarized in Table I. These scores were chosen because they reflect cognitive or mobility health and exhibited a sufficient variance (>0.1 after normalization to a $[0, 1]$ range) among the participants. The first such score is provided by the Wechsler Test of Adult Reading (WTAR) [30]. The WTAR estimates intellectual functioning, IQ, by having participants read aloud a list of 50 words with irregular pronunciations. This assesses knowledge and reading skills that are acquired over time as the test does not rely on standard pronunciation rules. The Telephone Interview of Cognitive Status (TICS) [22] offers a cognitive screening measure that can be linked directly to the widely-used Mini-Mental State Examination (MMSE). Unlike the MMSE, the TICS can be administered over the phone. In this test, cognitive function is assessed by having participants complete an 11-item brief screening test over the phone by completing tasks such as “count backwards by 7 from 100”. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [31] characterizes abnormal cognitive decline. Participants complete 12 subsets that assess the cognitive domains of attention, language, visuospatial/constructional abilities, and delayed memory. The F-A-S (FAS) [32] measures executive function. Participants are given one minute to list as many words as they can that begin with the letters “F”, “A”, or “S”. The Timed Up and Go (TUG) Test [33] is a widely-employed method of evaluating basic mobility maneuvers and has also been shown to offer indicators of cognitive and functional abilities [34].

We also include two measures of functional abilities. These exhibit lower variance among the normalized values (0.01 for DEX, 0.07 for ADCS-ADL) among participants. However, they provide insight on how behavior patterns reflect standardized clinical assessment methods and vice versa. The

first of these tests is the Dysexecutive Functioning Questionnaire (DEX) [35]. DEX asks participants to rate 20 items that are designed to assess everyday manifestations of difficulties consistent with common executive difficulties. These are emotion and personality changes, motivational changes, behavioral changes, and cognitive changes. Each item is scored on a 5-point Likert scale. The last targeted clinical score is generated by the Alzheimer’s disease Cooperative Study Instrumental protocol. In this inventory, ADCS-ADL [36], participants rate their ability (ranging from full independence to full dependence) to complete 6 basic activities of daily living (e.g., eating) and 16 instrumental activities of daily living (e.g., cooking) over the prior 4-week period. The remaining clinical parameters are used as supplemental clinical variables in the analysis. Fig. 2 shows the distribution of the target clinical scores projected using Principal Components Analysis onto a 4D space.

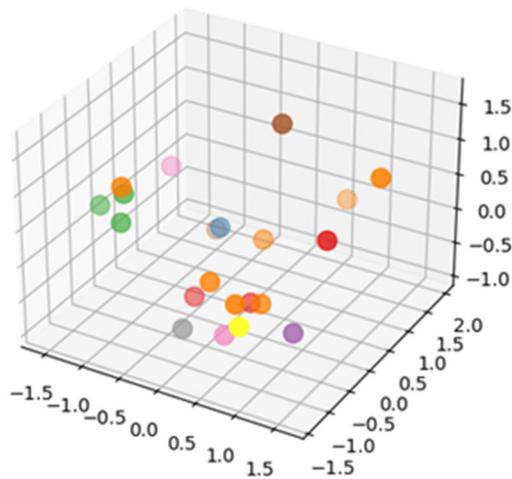


FIGURE 2. Projection of clinical variables onto a 4D space (3D geometric space and color). The project illustrates the distribution and variance of collected clinical information for the participant sample.

B. ACTIVITY RECOGNITION

We define a behaviorome as a set of digital behavior markers. The markers are more comprehensive if the set includes details of time spent on activities that are common and often included in clinical assessments. To include this information, we need a method that will automatically label sensor data with corresponding activity labels.

We have previously designed machine learning algorithms that label each ambient or wearable sensor reading with a corresponding activity label, referred to as Mobile-AR and Home-AR [37]¹. Researchers have designed numerous methods for human activity recognition from smart home sensors [37]–[40] and mobile sensors [41]–[46]. However, most methods operate in controlled laboratory conditions and no previously-reported work has investigated labeling and

modeling activities from both ambient and wearable sensor data. The Mobile-AR and Home-AR algorithms, in contrast, label activities in real-time for data collected in noisy, complex, realistic settings. Both algorithms employ a random forest with 100 trees to map a window of sensor readings to a corresponding activity label. The window size for ambient sensor readings is 30 readings (they do not arrive at equally-spaced time intervals) and for wearable sensor readings is 5 seconds. These window sizes have demonstrated success for activity recognition in earlier studies [25], [37], [47].

TABLE II
EXTRACTED FEATURES FOR ONE WINDOW OF SENSOR DATA.

<i>Ambient Sensors</i>	<i>Mobile Sensors</i>
time of day (hour, seconds past midnight), day of the week	max, min, sum, mean, median, mean/median abs value, standard deviation, mean/median abs deviation of sensor values, moments, mean and variance of Fourier Transformed sensor values
window duration	coefficient of variation, skewness, kurtosis, signal energy, log signal energy, power
time elapsed since previous sensor reading	sensor value autocorrelation
dominant (most common) sensor in current / previous window	correlations between axes for multidimensional sensors
last sensor / location to generate reading, location of last motion sensor	absolute value of successive values, time between peaks
window complexity (entropy of numbers of readings for each sensor)	heading change rate, stop rate, overall trajectory
number of changes in location	location type
number of sensors generating readings	normalized distance / direction from person’s mean location, distance traveled
number of readings for each sensor	location within most-frequented locations by time of day
time elapsed for each sensor since previous reading	
<i>Activities: Bed-toilet transition, Cook, Eat, Enter home, Leave home, Personal hygiene, Relax, Sleep, Wash dishes, Work</i>	<i>Activities: Chores, Eat, Entertainment, Errands, Exercise, Hobby, Hygiene, Relax, School, Sleep, Travel, Work</i>

¹ Code for Mobile-AR and Home-AR is available at <https://github.com/WSU-CASAS/AL> and <https://github.com/WSU-CASAS/AL-Smarthome>.

Each ambient sensor is identified based on its location and its type. Sensor locations are defined by regions of the home (bathroom, bedroom, dining room, entry, hall, kitchen, living room, office, other) and type (motion, door, temperature, light). These designators allow us to build models that generalize to new home settings without a need for floorplan information or a way to map sensor identifiers in one home to sensor identifiers in other homes. In the case of wearable data, Mobile-AR applies a Butterworth low-pass filter with a cutoff of 0.3Hz to sensor readings before extracting features to remove signal noise. The features extracted for each sensor modality are listed in Table II, together with the activity categories that are automatically recognized for each modality.

As shown in Table II, the learned activities are distinct for the different sensor platforms. Activity models are created from prior studies in which ground truth labels were provided for training. Activity categories were selected in those cases based on the types of activities that could be readily detected by the corresponding type of sensor. Many of the activity labels represent pre-defined activities of daily living (ADLs), while other labels are based on common behavior patterns.

The participants in the study we report here did not provide any ground truth activity labels for each smart home data or smartwatch data. As a result, we rely on pre-trained models to accurately provide activity information. Because the accuracy of the labels will impact the reliability of the extracted behaviorome, we evaluated the pre-trained models for their predictive performance on the activities listed in Table II.

TABLE III
HOME-AR ACTIVITY RECOGNITION RESULTS BASED ON LEAVE-ONE-HOME-OUT TESTING. PERFORMANCE METRICS ARE AVERAGE PRECISION, RECALL, AND F1-SCORE.

<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
0.86	0.85	0.85

TABLE IV
MOBILE-AR ACTIVITY RECOGNITION RESULTS BASED ON LEAVE-ONE-SUBJECT-OUT TESTING. PERFORMANCE METRICS ARE AVERAGE PRECISION, RECALL, AND F1-SCORE.

<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
0.94	0.87	0.86

We started by performing a leave-one-house-out performance analysis for Home-AR. The evaluation is based on 94 homes containing a total of 22,602,331 sensor reading sequences with corresponding ground truth activity labels. Table III summarizes the performance results for Home-AR. Similarly, we evaluate Mobile-AR using leave-one-subject-out evaluation for 250 individuals who provided activity labels for a total of 1,485,438 sensor reading sequences. Table IV summarizes the performance results for Mobile-AR. For both platforms, activity recognition performs at an f1 score of 0.85 or higher. This provides some evidence supporting inclusion of activity labels in the behaviorome that we describe in the next section.

C. CREATING A BEHAVIOROME

Using the activity-labeled time series data, we compute and compile digital behavior markers, which become the person's behaviorome. The set of digital behavior markers includes a set of behavior markers for each day and each hour of the day to describe observed behavior within that time period. From this information, we then additionally extract a set of overall behavior markers².

The daily, hourly, and overall behavior markers are summarized in Table V. As this table indicates, statistical summaries are gathered for each separate (daily, hourly) time period as well as for the entire data collection time period. In the case of smartwatch data, additional sensor data were available that were not labeled with the primary activity categories listed in Table II. For some earlier studies, participants were asked to provide any appropriate activity label. Many of these diverse labels only appeared for a few participants, represented specialized activities, or indicated a person's location context rather than an activity category. However, these extra labels can be valuable in building a behaviorome. Therefore, we train additional one-class random forest classifiers to recognize whether a person's current context belongs to one of these categories. For all learning tasks, the class distributions are imbalanced. To address this challenge, we weight each training point. Weights are inversely proportionate to the corresponding class relative frequency in the training data. As shown in Table V, the set of behavior markers includes time spent in one of these specialized activities. The one-class categories are: airplane, art, bathe, beach, biking, bus, car, chores, church, computer, cook, dress, drink, entertainment, groom, hobby, lunch, movie, music, relax, restaurant, school, service, shop, socialize, and sport.

² Code to extract digital behavior markers is available at <https://github.com/WSU-CASAS/DM>.

TABLE V
SUMMARY OF DIGITAL BEHAVIOR MARKERS THAT COMPRISE A PERSON'S BEHAVIOROME. MARKERS ARE CREATED SEPARATELY FOR SMART HOME DATA AND SMARTWATCH DATA, THEN FUSED INTO ONE BEHAVIOROME. IN THE CASE OF BOTH SMART HOME AND SMARTWATCH DATA, MARKERS ARE DEFINED FOR EACH DAY, EACH HOUR, AND OVERALL.

Modality	Time Period	Markers	Modality	Time Period	Markers
home	day	number of sensor readings number of distinct activities performed number of distinct locations visited time (seconds) spent on each activity time (seconds) spent at each location time of day (seconds past midnight) for first occurrence of each activity time of day (seconds past midnight) for first visit to each location	watch	day	total acceleration total rotation number of missing readings total distance traveled time (seconds) spent on each one-class activity and each primary activity time spent at each location type time of day (seconds past midnight) for first occurrence of each primary activity time of day (seconds past midnight) for first visit to each location type
	hour	number of sensor readings number of distinct activities performed number of distinct locations visited time spent on each activity time spent at each location		hour	total acceleration total rotation number of missing readings time spent on each one-class activity time spent on each primary activity time spent at each location type
	overall	statistics for daily behavior markers: mean, median, standard deviation, max, min, zero/mean crossings, interquartile range, skewness, kurtosis, signal energy statistics for hourly behavior markers: mean, median, standard deviation, max, min, zero/mean crossings, interquartile range, skewness, kurtosis, signal energy regularity index based on hourly values for number of sensor readings: within weeks, within weekdays, between weeks regularity index based on hourly values for number of activities performed: within weeks, within weekdays, between weeks regularity index based on hourly values for number of locations visited: within weeks, within weekdays, between weeks circadian rhythm: sensor reading count, number of activities, locations visited		overall	statistics for daily behavior markers: mean, median, standard deviation, max, min, zero/mean crossings, interquartile range, skewness, kurtosis, signal energy statistics for hourly behavior markers: mean, median, standard deviation, max, min, zero/mean crossings, interquartile range, skewness, kurtosis, signal energy regularity index of total acceleration: within weeks, within weekdays, between weeks regularity index of total rotation: within weeks, within weekdays, between weeks regularity index of total distance: within weeks, within weekdays, between weeks circadian rhythm: total acceleration, total rotation, total distance

Furthermore, the overall behavior markers contain information about a person's behavior routine as a whole, as observed over the entire data collection period. These include computation of a set of regularity indicators and circadian rhythm values. A regularity index is calculated using the hourly data and compares the uniformity of a person's schedule. This index is calculated for a within-week measure to compare days with a week, within-weekday measure to specifically determine the regularity of a Monday-through-Friday schedule, and between weeks to determine if there is a longer-term uniformity of schedule. The indices are calculated based on total acceleration, total rotation, and total distance traveled values. To calculate a regularity index, the data are first scaled to the range [-0.5, 0.5]. As introduced by Wang et al. [5], the regularity index comparing days a and b is defined as:

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

For our analysis, $T=24$ hours and x_t^a represents the scaled value for hour t of day a .

Finally, we compute values that reflect the circadian rhythm of a person's routine. The computation is applied only to the markers extracted for hourly time periods and is computed based on total acceleration, total rotation, and distanced traveled during the corresponding time period. To quantify circadian rhythm strength, we estimate a power spectral density using a periodogram. Based on Fourier analysis, the spectral energy of each frequency can be calculated using 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, resulting in a periodogram. This

periodogram generates values to a set of possible cycle lengths. The computed circadian rhythm is then represented by the normalized periodogram-derived value for a 24-hour cycle [48].

Fig. 3 illustrates a periodogram that was constructed using sensor data collected from one smart home (the graph contains separate plots for each week of data collection). In the case of this home, the sensor data most strongly support a 24-hour cycle length. In the next section, we will examine whether a behaviorome constructed from these features can predict the clinical state of participants in our study.

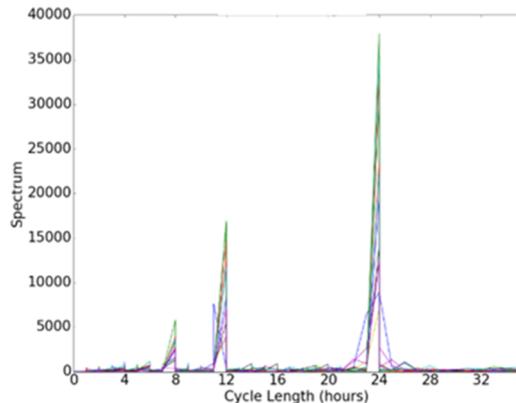


FIGURE 3. Circadian cycle strengths gleaned using a periodogram. The graph plots strengths of cycle sizes ranging from 0 to 34 hours.

D. JOINT INFERENCE OF CLINICAL SCORES

The goal of this work is to predict clinical scores from a sensor-derived behaviorome. We hypothesize that these digital behavior markers provide an indication of a person’s health status and thus offer the ability to predict clinical assessment scores. We further postulate that the relationship between multiple assessment measures that are typically collected can boost the predictive performance, by leveraging the relationship.

Fig. 4 illustrates the joint inference process. The intuition is that demographic and sensor information is used to infer values for each clinical parameter. In a second step, demographic and sensor information is combined with the inferred clinical parameters values (from step one) to predict values for each clinical parameter. If our hypothesis is valid, the inferred values from step two (joint inference) will result in greater predictive accuracy than the inferred values from step one (independent inference).

The independent predictor is our baseline clinical score predictor. As the name suggests, this predictor ignores the relational structure of the scores, and makes predictions using only information from the demographics and behaviorome. Here, the term “demographic” is broadly used to indicate the population sector and the health background of the person (e.g., substance abuse, high blood pressure). No clinical variables are used to predict values for other clinical variables.

More formally, the independent predictor is trained as follows. For each subject, we collect demographic features ψ_d , extract the digital behavior features ψ_{sh} and ψ_{sw} listed

in Table V, and gather ground truth scores for the target clinical scores ψ_{tc} listed in Table I as well as supplemental clinical scores ψ_{sc} described in Section IIIA. We train an independent regressor Π_i for each target and supplemental clinical score i . To do this, we collect the combined features as input x_i and the ground truth predictions as output y_i . We feed the aggregate set of input/output pairs, $\{x_i, y_i\}_{i=1}^N$ as training examples to a regression learner that seeks to minimize loss function L . For this problem, the non-negative loss function $L(x, \hat{y}) \in \mathbb{R}^+$ is the loss associated with labeling a particular input $x \in \mathbb{R}^+$ as output $\hat{y} \in \mathbb{R}^I$ when the true output is $y \in \mathbb{R}^I$ (normalized mean absolute error).

Our inference goal is to return a function/predictor whose outputs have low expected loss, or correspondingly yield high predictive performance. Random forest performs well for complex problems, such as Home-AR and Mobile-AR. However, the clinical score prediction problem exhibits even higher dimensionality. The independent predictor utilizes 1,038 features (86 demographic features and 952 behavior features). The joint predictor utilizes these 1,038 features plus 7 target clinical scores and 129 supplemental clinical scores. For very high-dimensional spaces, gradient boosting has demonstrated consistent success [49]. Using gradient boosting with 100 estimators and a learning rate of 0.1, multiple training iterations are performed during which additional regression trees are added to an ensemble. Each added tree reduces the least squares loss from the previous iteration. We incorporate gradient boosting with regression trees for the independent and joint prediction steps of the process shown in Fig. 4.

During the joint prediction step, we employ joint predictors to improve predictive accuracy. These joint predictors are pseudo-independent in the sense that each regressor predicts the output for a single clinical score but has independent-predictor outputs from all of the other regressors available as part of the input feature vector to make a more-informed prediction. The feature vector to predict clinical score i now consists of demographic features ψ_d , behaviorome features ψ_{sh} and ψ_{sw} , inferred supplemental clinical scores $\hat{\psi}_{sc}$, and inferred target clinical scores $\hat{\psi}_{tc}$ for all variables other than i . Joint prediction offers advantages when the variables being predicted are related – one score may be easier to predict than another, and its inferred value can improve prediction for the related variable.

IV. EXPERIMENTAL RESULTS

We evaluate the ability of the machine learning regression algorithm to predict clinical scores from the extracted behaviorome by computing Pearson’s correlation between the predicted and actual values. We compare the results using smart home behavior markers, smartwatch behavior markers, and a fusion of the two. Additionally, we compare the results generated by independent predictors with those generated by joint predictions. All results are generated from a leave-one-subject-out validation. Results are summarized in Table VI.

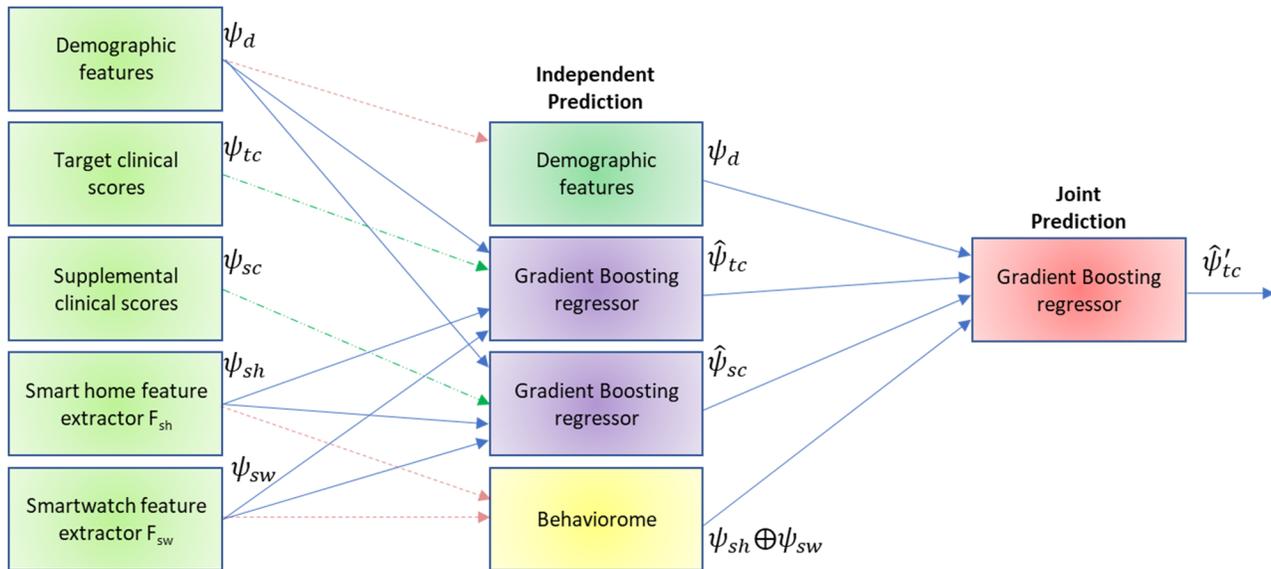


FIGURE 4. The joint inference process. Demographic variables and digital behavior markers are initially used to predict values for the target clinical scores and supplemental clinical scores, in an *independent inference* step. During *joint inference*, the predicted clinical scores for all variables except i are combined with demographic features and the behaviorome to predict the value of clinical score i . Dotted lines indicate variables that are used for both rounds. Dashed lines indicate features that are used as ground truth only during training.

TABLE VI

CORRELATION BETWEEN PREDICTED AND ACTUAL CLINICAL SCORES. RESULTS ARE REPORTED USING SMART HOME BEHAVIOR MARKERS, SMARTWATCH BEHAVIOR MARKERS, AND A FUSION OF MARKERS FROM BOTH SENSOR MODALITIES. RESULTS ARE REPORTED FOR EACH TARGET CLINICAL SCORE BASED ON INDEPENDENT PREDICTION (I) AND JOINT PREDICTION (J).

Measure	SmartHome-I	Smartwatch-I	Fusion-I	SmartHome-J	Smartwatch-J	Fusion-J
WTAR	0.151	0.323	0.236	0.856*	0.879*	0.848*
RBANS	0.337	0.324	0.358	0.928*	0.962*	0.932*
TICS	0.180	0.358	0.354	0.019	0.035	0.188
FAS	0.238	0.278	0.498	0.806*	0.708*	0.764*
TUG	0.301	0.245	0.353	0.227	0.492	0.278
DEX	0.617	0.202	0.063	0.866*	0.881*	0.451*
ADLC	0.279	0.249	0.303	0.437*	0.391	0.252
Average	0.301	0.283	0.309	0.591	0.621	0.590
Average of independent predictions = 0.298				Average of joint predictions = 0.601		

* The result is significant at $p < 0.05$.

Correlations ranging from $r=0.019$ (TICS, joint prediction based on smart home behavior markers) to $r=0.962$ (RBANS, joint prediction based on smartwatch features) result from the proposed prediction method. The highest-performing prediction strategy for each clinical score is a joint inference method. In fact, the average correlation over all behavior marker combinations for independent predictors is a small correlation ($r=0.298$) while the average for joint predictors is a moderate correlation ($r=0.601$).

The results provide evidence that digital behavior markers from both smart home sensors and smartwatch sensors are predictive of clinical scores. The sensor modality that provides the largest predictive performance varies by

category. Interestingly, a fusion of markers from both the smart home and the smartwatch does not consistently yield the highest predictive performance. While gradient boosting performs better than a random forest regressor and a regression tree classifier (which average $r=0.592$ and $r=0.499$, respectively, for joint prediction based on a fusion of behavior markers), the learner still appears to struggle with the high-dimensional space. Dimensionality reduction techniques can be further explored in future work, although reducing the dimensionality with PCA (number of components = 21) reduces the predictive performance to $r=0.259$.

V. CONCLUSIONS

In this study, we assess the ability of multi-modal sensor data to provide digital behavior markers that are predictive of clinical health states. To do this, we describe a method of collecting sensor data, automatically labeling the data with activity labels, and extracting features that form a person's behaviorome. We built a behaviorome for each person in our study from a combination of smart home and smartwatch data. Using a gradient boosting regression algorithm, the method predicts clinical scores with moderate or high correlation for WTAR, RBANS, FAS, TUG. These scores include measures of cognitive and mobility-based health. The model also predicts scores with moderate correlation for DEX and ADLC, which are self-report measures of independent functioning. We further provide evidence that performing joint prediction for a large set of clinically-related variables boosts predictive performance.

Limitations of this work include the small sample size of 21 healthy older adults. Future studies will involve participants exhibiting a range of health conditions to examine the ability to further differentiate between diagnosis groups. Continued analysis may also shed light on the types of sensor modalities that are most predictive of specific clinical scores. We will also adapt the approach described here to track and predict changes in health status over time.

ACKNOWLEDGMENT

The authors would like to thank members of the WSU CASAS lab for their assistance in this study. We would also like to thank Bryan Minor and Brian Thomas for their help in designing data collection tools. We also thank members of the WSU Neuropsychology and Aging laboratory for their work collecting and scoring the clinical data.

REFERENCES

- [1] R. Tourangeau, L. J. Rips, and K. Rasinski, *The Psychology of Survey Response*. Cambridge University Press, 2000.
- [2] M. G. Palmer and C. M. Johnson, "Experimenter presence in human behavior analytic laboratory studies: Confound it?," *Behavior Analysis: Research and Practice*, vol. 19, no. 4, pp. 303–314, 2019.
- [3] M. Mancini *et al.*, "Continuous monitoring of turning mobility and its association to falls and cognitive function: A pilot study," *The Journals of Gerontology: Series A: Biological Sciences and Medical Sciences*, vol. 71, no. 8, pp. 1102–1108, 2016.
- [4] D. A. Adler *et al.*, "Predicting early warning signs of psychotic relapse from passive sensing data: An approach using encoder-decoder neural networks," *JMIR Mhealth Uhealth*, vol. 8, no. 8, p. e19962, 2020.
- [5] R. Wang *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.
- [6] J. F. Huckins, A. W. DaSilva, W. Want, E. L. Hedlund, and C. Rogers, "Mental health and behavior during the early phases of the COVID-19 pandemic: A longitudinal mobile smartphone and ecological momentary assessment study in college students," *Journal of Medical Internet Research*, vol. 6, p. e20185, 2020.
- [7] P. Newland *et al.*, "Associations between self-reported symptoms and gait parameters using in-home sensors in persons with multiple sclerosis," *Rehabilitation Nursing Journal*, vol. 45, no. 2, pp. 80–87, 2020.
- [8] L. Fiorini, M. Maselli, and E. Castro, "Feasibility study on the assessment of auditory sustained attention through walking motor parameters in mild cognitive impairments and healthy subjects," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2017, pp. 897–900.
- [9] R. Fallahzadeh, B. Minor, L. Evangelista, D. J. Cook, and H. Ghasemzadeh, "Mobile sensing to improve medication adherence," in *ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2017, pp. 279–280.
- [10] R. Ravichandran, S.-W. Sien, S. Patel, J. A. Kientz, and L. R. Pina, "Making sense of sleep sensors: How sleep sensing technologies support and undermine sleep health," in *Conference on Human Factors in Computing Systems*, 2017.
- [11] S. Fallmann and L. Chen, "Computational sleep behavior analysis: A survey," *IEEE Access*, vol. 7, p. 142421, 2019.
- [12] S. Lee, R. M. Walker, Y. Kim, and H. Lee, "Measurement of human walking movements by using a mobile health app: Motion sensor data analysis," *JMIR Mhealth Uhealth*, vol. 9, no. 3, p. e24194, 2021.
- [13] M. Lin and W.-J. Hsu, "Mining GPS data for mobility patterns: A survey," *Pervasive and Mobile Computing*, vol. 12, pp. 1–16, 2014.
- [14] M. Boukhechba, Y. Huang, P. Chow, K. Fua, B. A. Teachman, and L. E. Barnes, "Monitoring social anxiety from mobility and communication patterns," *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 749–753, 2017.
- [15] V. Tseng, J. Costa, M. Jung, and T. Choudhury, "Using smartphone sensor data to assess inhibitory control in the wild: Longitudinal study," *JMIR Mhealth Uhealth*, vol. 8, no. 12, p. e21703, 2020.
- [16] J. Austin, H. H. Dodge, T. Riley, P. G. Jacobs, S. Thielke, and J. Kaye, "A smart-home system to unobtrusively and continuously assess loneliness in older adults," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 4, 2016.
- [17] A. Akl, B. Taati, and A. Mihailidis, "Autonomous

- unobtrusive detection of mild cognitive impairment in older adults,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1383–1394, 2015.
- [18] P. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, “Longitudinal functional assessment of older adults using smart home sensor data,” *IEEE Journal of Biomedical and Health Informatics*, 2015.
- [19] A. Alberdi Aramendi *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.
- [20] R. Fritz, M. Wilson, D. Dermody, M. Schmitter-Edgecombe, and D. J. Cook, “Automated smart home assessment to support pain management,” *Journal of the American Medical Informatics Association*, 2020.
- [21] M. Skubic, G. Alexander, M. Popescu, M. Rantz, and J. Keller, “A smart home application to eldercare: Current status and lessons learned,” *Technology and Health Care*, vol. 17, no. 3, pp. 183–201, 2009.
- [22] T. G. Fong *et al.*, “The Telephone Interview for Cognitive Status: Creating a crosswalk with the Mini-Mental State Exam,” *Alzheimer’s and Dementia*, vol. 5, no. 6, pp. 492–497, 2009.
- [23] D. J. Cook, A. Crandall, B. Thomas, and N. Krishnan, “CASAS: A smart home in a box,” *IEEE Computer*, vol. 46, no. 7, pp. 62–69, 2012.
- [24] D. Liciotti, M. Bernardini, L. Romeo, and E. Frontoni, “A sequential deep learning application for recognising human activities in smart homes,” *Neurocomputing*, vol. 396, pp. 501–503, 2020.
- [25] D. J. Cook, N. Krishnan, and P. Rashidi, “Activity discovery and activity recognition: A new partnership,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 43, no. 3, pp. 820–828, 2013.
- [26] Y. Asim, M. A. Azam, M. Ehatisham-ul-Haq, U. Naeem, and A. Khalid, “Context-aware human activity recognition (CAHAR) in-the-wild using smartphone accelerometer,” *IEEE Sensors*, vol. 8, pp. 4361–4371, 2020.
- [27] R. Ramamurthy and N. Roy, “Recent trends in machine learning for human activity recognition—A survey,” *Data Mining and Knowledge Discovery*, 2018.
- [28] M. Kheirhahan *et al.*, “A smartwatch-based framework for real-time and online assessment and mobility monitoring,” *Journal of Biomedical Informatics*, vol. 89, pp. 29–40, 2019.
- [29] M. Schmitter-Edgecombe, C. A. Sumida, and D. J. Cook, “Bridging the gap between performance-based assessment and self-reported everyday functioning: An ecological momentary assessment approach,” *The Clinical Neuropsychologist*, 2020.
- [30] The Psychological Corporation, *Wechsler Test of Adult Reading*. San Antonio, TX: Harcourt Assessment, 2001.
- [31] C. Randolph, M. C. Tierney, D. Mohr, and T. N. Chase, “The repeatable battery for the assessment of neuropsychological status (RBANS): Preliminary clinical validity,” *Journal of Clinical and Experimental Neuropsychology*, vol. 20, no. 3, pp. 310–319, 1998.
- [32] O. Spreen and A. L. Benton, “Neurosensory center comprehensive examination for aphasia,” 1977.
- [33] D. Podsiadlo and S. Richardson, “The timed ‘Up and Go’ test: A test of basic functional mobility for frail elderly persons,” *Journal of American Geriatric Society*, vol. 39, pp. 142–148, 1991.
- [34] G. Sprint, D. Cook, and D. Weeks, “Towards automating clinical assessments: A survey of the Timed Up and Go (TUG),” *Biomedical Engineering, IEEE Reviews in*, vol. PP, no. 8, pp. 64–77, 2015, doi: 10.1109/RBME.2015.2390646.
- [35] D. Gerstorf, K. L. Siedlecki, E. M. Tucker-Drob, and T. A. Salthouse, “Executive dysfunctions across adulthood: Measurement properties and correlates of the DEX self-report questionnaire,” *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, vol. 15, no. 4, pp. 424–445, 2008.
- [36] D. Galasko *et al.*, “An inventory to assess activities of daily living for clinical trials in Alzheimer’s disease. The Alzheimer’s Disease Cooperative Study,” *Alzheimer Disease and Associated Disorders*, vol. 11 (Supple, pp. S33–S39, 1997.
- [37] S. Aminikhanghahi and D. J. Cook, “Enhancing activity recognition using CPD-based activity segmentation,” *Pervasive and Mobile Computing*, vol. 53, no. 75–89, 2019.
- [38] B. Minor, J. R. Doppa, and D. J. Cook, “Learning activity predictors from sensor data: Algorithms, evaluation, and applications,” *IEEE Transactions on Knowledge and Data Engineering*, 2017.
- [39] J. Wan, M. Li, M. O’Grady, X. Gu, M. Alawlaqi, and G. O’Hare, “Time-bounded activity recognition for ambient assisted living,” *IEEE Transactions on Emerging Topics in Computing*, 2018.
- [40] Y. Du, Y. Lim, and Y. Tan, “A novel human activity recognition and prediction in smart home based on interaction,” *Sensors*, vol. 19, p. 4474, 2019.
- [41] P. Bharti, D. De, S. Chellappan, and S. K. Das, “HuMAN: Complex activity recognition with multi-modal multi-positional body sensing,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 4, pp. 857–870, 2019.
- [42] M.-C. Kwon, H. You, J. Kim, and S. Choi, “Classification of various daily activities using convolution neural network and smartwatch,” in *IEEE International Conference on Big Data*, 2018.

- [43] G. Wilson, J. Doppa, and D. J. Cook, “CoDATS: A convolutional deep domain adaptation model for time series sensor data,” in *ACM Conference on Knowledge Discovery and Data Mining*, 2020.
- [44] K. D. Feuz and D. J. Cook, “Modeling skewed class distributions by reshaping the concept space,” in *AAAI Conference on Artificial Intelligence*, 2017, pp. 1891–1897.
- [45] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, “Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges,” *Expert Systems with Applications*, vol. 105, pp. 233–261, 2018.
- [46] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.
- [47] C. Culman, S. Aminikhangahi, and D. J. Cook, “Easing power consumption of wearable activity monitoring with change point detection,” *Sensors*, vol. 20, no. 1, p. 310, 2020.
- [48] J. Williams and D. Cook, “Using time series techniques to forecast and analyze wake and sleep behavior,” in *KDD Workshop on Mining and Learning from Time Series*, 2016.
- [49] R. Blagus and L. Lusa, “Gradient boosting for high-dimensional prediction of rare events,” *Computational Statistics & Data Analysis*, vol. 113, pp. 19–37, 2017.



DIANE J. COOK received the B.S. degree in math / computer science from Wheaton College, Wheaton, IL, USA in 1985 and the M.S. and Ph.D. degrees in computer science from the University of Illinois, Urbana, IL, USA in 1987 and 1990, respectively. She became a Member (M) of IEEE in 1995 and a Fellow (F) in 1997. She is a Regents Professor and Huie-Rogers Chair in the School of Electrical Engineering and Computer Science at Washington State University. She is an Associate Editor of the journals *Proceedings of the IEEE*, *IEEE Journal of Biomedical and Health Informatics*, *ACM Transactions on Intelligent Systems and Technology*, *ACM Transactions on Knowledge Discovery from Data*, *IEEE Transactions on Knowledge and Data Engineering*, and *ACM/IMS Transactions on Data Science*.



MAUREEN SCHMITTER-EDGECOMBE received the B.S. in Biology from Bucknell University in Lewisburg, PA, USA in 1988 and the M.S. and Ph.D. degrees in clinical neuropsychology from the University of Memphis in Memphis, TN, USA in 1991 and 1994, respectively. She is a Professor in the Department of Psychology at Washington State University. She is an Associate Editor of the journals *Neuropsychology*, *Journal of the International Neuropsychological Society*, *Journal of Clinical and Experimental Neuropsychology*, and *The Clinical Neuropsychologist*.