# HuMAn: Complex Activity Recognition with Multi-Modal Multi-Positional Body Sensing

Pratool Bharti, Debraj De, Sriram Chellappan, and Sajal K. Das, *Fellow, IEEE*

**Abstract**—Current state-of-the-art systems in the literature using wearables are not capable of distinguishing a large number of fine-grained and/or complex human activities, which may appear similar but with vital differences in context, such as lying on floor versus lying on bed versus lying on sofa. This paper fills the gap by proposing a novel system, called *HuMAn*, that recognizes and classifies complex at-home activities of humans with wearable sensing. Specifically, *HuMAn* makes such classifications feasible by leveraging selective multi-modal sensor suites from wearable devices, and enhances the richness of sensed information for activity classification by carefully leveraging placement of the wearable devices across multiple positions on the human body. The *HuMAn* system consists of the following components: (a) a practical feature set extraction method from selected multi-modal sensor suites; and (b) a novel two-level structured classification algorithm that improves accuracy by leveraging sensors in multiple body positions; and (c) improved refinement in classification of complex activities with minimal external infrastructure support (e.g., only a few Bluetooth beacons used for location context). The proposed system is evaluated with 10 users in real home environments. Experimental results demonstrate that the *HuMAn* system can detect 21 complex at-home activities with high degree of accuracy. For same-user evaluation strategy, the average activity classification accuracy is as high as 95 percent over all of the 21 activities. For the case of 10-fold cross-validation evaluation strategy, the average classification accuracy is 92 percent, and for the case of leave-one-out cross-validation strategy, the average classification accuracy is 75 percent.

**Index Terms**—Complex activity recognition, smart health, smartphone multi-modal sensors, conditional random fields

## 1 Introduction

Daily activities of people are complex, and consist of one or more than one unit-level sub-activities [43]. Automated classification of human activity contexts (ranging from simple activities to more complex ones) are important for applications like smart healthcare [7], [41], quantified self [48], monitoring elderly people in assisted living [10], designing smart homes and appliances [33], activity-aware media content delivery [34], and so on. Although a significant body of literature exists for activity context recognition, some of them incur high infrastructure costs or direct privacy concerns, and above all, the majority of existing works are able to recognize mostly coarse-grained Activities of Daily Living (ADL) and only very few complex Instrumental Activities of Daily Living (IADL) [49]. Coarse-grained ADLs are typically basic self-care skills that people learn during early childhood, such as sitting, standing, walking, watching TV, etc. whereas IADLs are complex tasks needed for independent living (usually learnt later) such as cooking, housekeeping, doing laundry, etc. In essence, basic ADLs often include more physical or postural activities, while IADLs require a combination of physical and cognitive efficiencies. Recognition of complex activities in humans is a challenging problem, requiring innovative research solutions. This motivates our work.

### 1.1 Contributions of this Paper

In this paper, we design a novel system called *HuMAn*, which stands for *Hybrid Multi-modal and body multi-positional system for complex Activity recognition*. The overall architecture of the *HuMAn* system is illustrated in Fig. 1 that can recognize 21 complex activities.

Our system significantly improves quality (in terms of the accuracy) as well as quantity (in terms of the number) of activities detected as compared to the existing works. This is due to a combination of three factors: (i) multi-modal sensing, (ii) context awareness from sensors placed at multiple body positions, and (iii) location awareness using simple Bluetooth beacons. These are described below.

*Multi-modal Sensing.* Current smartphones and smart wearable devices are equipped with versatile sensing capabilities with multi-modal sensor arrays. These technological developments are critical to classify complex activities of interest in this paper. As an example, sensory data from an accelerometer may indicate that a person is sitting, but when combined with the data from a humidity sensor, one may infer that the person is sitting in a bathroom. Also, when combined with an altitude sensor, one may even infer which floor the person is currently at. Similarly, a gyroscope can indicate that the person's wrist is moving, but when integrated with a temperature sensor, we could glean that the person is in a kitchen, which can help better differentiate between activities like cooking, cleaning utensils, or opening a fridge. Examples like these and more make up the design premise of our *HuMAn* system.

Specifically, *HuMAn* utilizes a careful combination of existing sensor suites in smartphones to detect several body

- *P. Bharti and S. Chellappan are with the Computer Science & Engineering Department, University of South Florida, Tampa, FL 33620. E-mail: pratool@mail.usf.edu, sriramc@usf.edu.*
- *D. De is with the Smart Health Beacons LLC, Rolla, MO 6540. E-mail: debraj.de1@gmail.com.*
- *S. K. Das is with the Computer Science Department, Missouri University of Science and Technology, Rolla, MO 65409. E-mail: sdas@mst.edu.*

*Manuscript received 9 Feb. 2016; revised 2 Apr. 2018; accepted 7 May 2018. Date of publication 7 June 2018; date of current version 4 Mar. 2019. (Corresponding author: Debraj De.)*
*For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.*
*Digital Object Identifier no. 10.1109/TMC.2018.2841905*

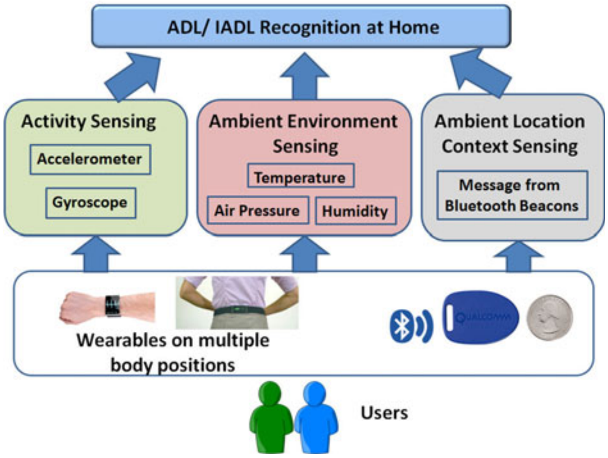

ADL/ IADL Recognition at Home
Activity Sensing
Accelerometer
Gyroscope
Ambient Environment Sensing
Temperature
Air Pressure
Humidity
Ambient Location Context Sensing
Message from Bluetooth Beacons
Wearables on multiple body positions
Users

Fig. 1. The *HuMAn* system for at-home ADL/IADL recognition system with multi-modal and body multi-positional wearable sensing.

locomotion activities (via accelerometer and gyroscope). Then, the context of such activities are refined by sensing sensors the ambient environment (e.g., temperature or humidity sensors) and relative altitude (e.g., a barometric air pressure sensor). To the best of our knowledge, there exists very little work that uses multi-modal ambience sensing for complex activity recognition. The work in [13] used atmospheric pressure sensing, but only for differentiating whether the user is present at indoor or outdoor. The work in [37] utilizes humidity sensor (along with other sensors like audio and bio-sensing), but mainly for detecting different outdoor sports and social activities of users. The key challenges in this realm are identifying the right modalities and features for different activities, demonstrating the feasibility of multi-modal integration for specific activities classified, and actually fusing the sensory data at run-time despite differences in sampling rates and sensitivities. We addressed these challenges in our *HuMAn* system.

*Context Awareness from Multiple Body Positions*. The second novelty of the *HuMAn* system is the extraction of additional contextual information from sensor suites by placing them on multiple positions on the human body. For instance, an accelerometer on the leg can indicate that a person is standing, but when combined with a gyroscope sensor on the wrist, one can differentiate between simply standing, or standing while talking, cooking or cleaning utensils. Similar challenges also lie when fusing multi-modal sensors from multiple positions in the body, which are addressed in this paper.

Specifically, we demonstrate how devices placed in different positions of human body (e.g., waist, back, thigh, wrist) provide subtle, but distinct signatures on activities by themselves. This, coupled with the sensed information from ambient environments, altitudes and beacon locations provide us with a superior set of features for much more accurate detection of complex human activities. Note that there are already a number of commercial products (e.g., ProeTEX [15]) on textile-based smart wearables that come integrated with embedded sensors at different positions within the textile for sensing information from various locations in the human body. Other such wearables include Lumo Back (on waist or lower back) [3], Lumo Lift (on back) [4], Nike+ (on legs or shoes) [5], Fitbit (on wrist) [2] and Biostrap (on wrist and feet) [1]. Therefore, our contributions in this paper are feasible with today's wearable technologies.

TABLE 1
List of Activities Detected by Our *HuMAn* System

| Activities | Abbreviation |
|---|---|
| Standing and Cleaning Utensils | CLNG_UTENSILS |
| Standing and Cooking | COOK |
| Standing and Leaning on Wall | LEAN_ON_WALL |
| Lying On Bed | LYNG_ON_BED |
| Lying On Floor | LYNG_ON_FLOOR |
| Lying On Sofa | LYNG_ON_SOFA |
| Running | RUN |
| Standing | STAND |
| Standing and Using Fridge | STAND_FRIDGE |
| Standing and Talking | STAND_TALK |
| Sitting and Eating | STNG_EATING |
| Sitting on Bed | STNG_ON_BED |
| Sitting on Commode | STNG_ON_COMM |
| Sitting on Floor | STNG_ON_FLOOR |
| Sitting on Sofa | STNG_ON_SOFA |
| Standing and Using Sink | USING_SINK |
| Walking | WALK |
| Walking Downstairs | WALK_DWNSTR |
| Walking Upstairs | WALK_UPSTR |
| Walking Indoor to Outdoor | IN_TO_OUT |
| Walking Outdoor to Indoor | OUT_TO_IN |

*Bluetooth Beacon to Leverage Location Context*. The indoor locations of a user at room level granularity are very useful to predict activities. For example, a set of activities performed in a bathroom is very different from the ones performed in a kitchen. For our study, we installed a few small and cheap Bluetooth beacons on walls across the home to get the subject's coarse location. Based on the beacon id and RSSI (Received Signal Strength Indicator), we approximate the subject's position in the home. Note that, while these location beacons provide room level granularity, the privacy concerns are far less than that for video cameras.

*Demonstrating the Classification of Complex Activities*. By conducting detailed experiments, we demonstrate that our *HuMAn* system can classify 21 complex human activities with high accuracy. These activities are listed in Table 1. To the best of our knowledge, this is the first time that 21 at-home activities are recognized via wearable devices. This is significantly higher compared to between 6-12 at-home activities reported in most existing works [12], [28], [53], [55]. Furthermore, *HuMAn* does not need expensive infrastructures like networks of sensors or cameras, which is an advantage of our system from cost and privacy perspectives.

## 1.2 Relevance of Activities to Healthcare

Upon discussions with healthcare experts in diverse areas, we see a relevance and need to classify the 21 complex activities in this paper. Caregivers will derive benefit if they are aware when patients with Dementia move from indoors to outdoors or cook repeatedly or walk upstairs/ downstairs too often. Even if not in real-time, progression of Dementia can be comprehended from detecting such activities over time. As another example, in the case of Hemorrhoids, a simple feedback message can be given to a person sitting for too long in a toilet. Activities related to eating, sitting for too long, and running are very important to monitor obesity and heart health. We point out that state-of-the-art work in ADL/ IADL recognition are limited in their ability to detect such complex activities, while our proposed *HuMAn* system can do so with very good accuracy for real-world healthcare applications. This is the novelty and impact of our proposed work.

### 1.3 Operation of *HuMAn* System

The proposed *HuMAn* system consists of three phases: (i) initial pre-computation with training data and feature selection, (ii) complex activity classification at each device, and (iii) integration of decisions from each device to classify final activity. In the *first phase*, the training dataset is used for feature extraction and also for training a model based on multi-scale Conditional Random Field (CRF) [31] based machine learning algorithm. Note that the same training dataset is also used for learning weights for each activity-device pair. These weights are essentially the precision (confidence) of a device at a specific body position while classifying a certain activity. The model and weights are then used for activity classification on the test dataset. The *second phase* of complex activity classification at each device works as follows. Initially the *feature set extraction* is done with the help of multiple sensor data sources from each wearable device. The devices in turn perform fine-grained activity classification using the learned CRF model. In our system, each wearable device is placed at a specific position in the body (i.e., waist, lower back, thigh and wrist) to gain contextual advantage of its placement. To utilize the processing power of each device and balancing the load, each device predicts an activity independently. In the *third and final phase*, the classified activity from each wearable is contextually integrated to predict one final activity.

We evaluated the *HuMAn* system in real home settings with 10 users, where for each user, smartphones were placed on the waist, lower back, thigh and wrist. Experimental results revealed that our system can detect 21 complex at-home activities with high accuracy. For same-user evaluations, the average activity classification accuracy is as high as 95 percent over all the 21 activities. . For the case of cross-user evaluations, the average classification accuracy is 92 and 75 percent, for 10-fold cross-validation and leave-one-out cross-validation evaluation, respectively. Note that preliminary results of this work were published in [16] and [46].

The rest of the paper is organized as follows. Section 2 reviews related work while Section 3 proposes the *HuMAn* system. The individual components of the *HuMAn* are presented in details in Sections 4, 5 and 6. Experimental evaluation and validation results are reported in Section 7. Finally, Section 8 concludes the paper with directions of future work.

## 2 Related Work

Important fundamental concepts, and a comprehensive survey of the literature on complex human activity recognition are discussed in [43] and [32]. There exist three main categories of works on activity recognition: (i) only with wearable devices [7], [47], [53], [54]; (ii) combining wearable devices and external static infrastructure based systems [42]; and (iii) with non-wearable technologies [17], [51], [50].

With only wearable devices, the activities can be classified by learning from data sensed by smartphones, wearable health tracker devices, smartwatches, near field communication (NFC) based gadgets, augmented reality devices (e.g., Google Glass), etc. For example, the work in [19] uses a belt-clip accelerometer tied to the waist to detect six activities, while another work in [25] leverages an accelerometer sensor in the subject's dominant wrist to classify seven human activities.

There also exists a body of literature that accomplishes activity recognition by combining data sensed from wearable devices and additional static infrastructures. As an example, the work in [42] integrates data sensed from infrared motion sensors mounted on ceilings of different rooms, where data are generated from a smartphone sensor in the pocket of a human user for classifying postural/ locomotive states of multiple humans inhabiting a home. Our prior work related to activity recognition is reported in [16]. We fused multi-sensor data from smartphones (used as wearables placed on multiple body positions) and Bluetooth beacons to classify 19 human activities. This prior work did not consider more complex activities related to ambient sensing (moving from indoors to outdoors, or from outdoors to indoors) like we do in the current paper. Furthermore, the feature selection techniques in our prior work [16] were primitive leading to lower accuracy of only 80 percent in activity classification for a single user. In contrast, the current paper uses superior feature extraction techniques, noise reduction, and parameter tuning to significantly improve accuracy of activity classification. Additionally, more thorough evaluations are done (e.g., cross-user 10-fold and leave-one-out cross-validation) considering the complexity of activities identified here to validate the performance of the *HuMAn* system. In another prior work [46], we investigated deep-learning approaches for human activity recognition by combining sensory data from smartphones and Bluetooth beacons, but the complexity of deep-learning algorithms on smartphones is a major limitation there.

Finally, there exist notable work in activity recognition with non-wearable devices. Using a combination of motion detectors, break-beam sensors, pressure mats, and contact switches, the work in [50] accomplishes motion tracking and limited activity recognition, such as sleeping. There also exist works where Microsoft Kinect RGB, infrared (IR) and 3D depth cameras were used for activity recognition [9], [17], [24], and [51]. Another work in [8] uses Active Sonar technologies for activity recognition and classification. The above techniques are not ubiquitous in the sense that the hardware used for classification are expensive, and as such, cannot cover an entire home. Typically, the hardware is fixed in a particular location (say in one room), and relocating them to other rooms as the subject moves away (from that room) is too cumbersome and not practical. Furthermore, in the case of cameras, privacy concerns are also raised. In [11], in-home WiFi signals are used for activity recognition. However, WiFi based techniques cannot distinguish activities which require no to minimal body movement (e.g., between 'sitting in sofa' versus 'sitting in bed'), or activities which are contextually very similar in location and body movement (e.g., between 'standing and cooking' versus 'standing and cleaning utensils'). This is because the change in Channel State Information (CSI) used for classification is very limited or similar in either case, which makes the classification very hard. Moreover, activity classification solutions based on WiFi signals may need recalibration based on the environment, hardware and associated signal strengths, which again can complicate practical adoption [52]. For energy management in wearables for activity recognition, kinetic energy harvesting techniques have been used to produce energy from user's motion [22], [26], and [27].

It is worth noting that the above works can classify about 6 to 12 human activities as shown in Table 1, while our *HuMAn* system can classify 21 activities. Table 2 provides a comprehensive comparison of related works with our *HuMAn* system.

TABLE 2
Comparison of Literature on Activity Classification

| **Activity classification work** | **Wearable sensors in use** | **Additional infrastructure in use** | **Activities recognized** |
|---|---|---|---|
| Gupta et al. [19] | Belt-clip Accelerometer | None | 6 activities: walking, jumping, running, sit to-stand/stand-to-sit, stand-to-kneel-to-stand, and being stationary |
| Kao et al. [25] | accelerometer | None | 7 activities: brushing teeth, hitting, knocking, working at a PC, running, swinging, walking |
| Riboni et al. [40] | Accelerometer, GPS | None | 10 activities: brushing teeth, hiking up, hiking down, riding bicycle, jogging, standing still, strolling, walking downstairs, walking upstairs, writing on blackboard |
| Zhu et al. [55] | Accelerometer, gyroscope, magnetometer, temperature | None | 11 activities: Standing, sitting, sleeping, sitting-to-standing, standing-to-sitting, level walking-to stair walking, stair walking-to-level walking, walking level, walking upstairs, walking downstairs, running |
| Cheng et al. [12] | Electrodes on neck, chest, leg and wrist | None | 11 activities: bread swallow, water swallow, chew, nod, shake head, look down, speak, look up, look left, look right, look straight |
| Khan et al. [28] | Accelerometer | None | 15 similar activities: standing, sitting, lying, lie-stand, stand-lie, sit-lie, lie-sit, sit-stand, stand-sit, walk-stand, stand-walk, walking, walking-upstairs, walking-downstairs, running |
| Zhan et al. [53] | Smartphone accelerometer and video cameras | None | 12 activities: walking, going upstairs, going downstairs, drinking, stand up, sit down, sitting, reading, watching TV/monitor, writing, switch water-tap, hand-washing |
| Roy et al. [42] | Smartphone accelerometer and gyroscope | Ceiling mounted infrared (IR) motion sensors | 6 “low-level” postural or motion activities: sitting, standing, walking, running, lying, climbing stairs 6 “high-level” semantic activities: cleaning, cooking, medication, sweeping, washing hands, watering plants |
| Wilson et al. [50] | None | Motion detectors, break-beam sensors, pressure mats, and contact switches | room-level tracking and basic activities such as sleeping on bed |
| Gaglio et al. [17] | None | Microsoft Kinect RGB and IR video camera | 10 gestures: horizontal arm wave, high arm wave, two hand wave, high throw, draw x, draw tick, forward kick, side kick, bend, clap hands 8 actions: catch cap, toss paper, take umbrella, walk, phone call, drink, sit down, stand up |
| Yang et al. [51] | None | 3D video camera | 16 Daily activities: drink, eat, read book, call cellphone, write, use laptop, vacuum clean, cheer up, sit still, toss paper, play game, lie down, walk, play guitar, stand up, and sit down |
| Blumrosen et al. [9] | None | Microsoft Kinect RGB and IR video camera | 2 activities: Walking in a complex pattern in relatively limited space, and repetitive hand tapping |
| Chen et al. [11] | None | In-home Wi-Fi | 6 activities: Pick up from the ground and stand up, sit down on a chair, stand up from a chair, lie down onto the mattress, stand up after lie down, fall |
| Blumrosen et al. [8] | None | Active sonar | 3 activities: Standing, walking and swinging arms |
| **Our proposed *HuMAn*system** | Wearable (body multi-positional) multi-modal sensors: accelerometer, gyroscope (for body locomotion); temperature, atmospheric pressure, humidity (for ambient environment); GPS, Bluetooth reception (for location context) | Bluetooth beacon in the physical environment | 21 fine-grained activity classes: (i) *Locomotive* (walk indoor, run indoor), (ii) *Semantic* (use refrigerator, clean utensil, cooking, sit and eat, use bathroom sink, stand and talk), (iii) *Transitional* (indoor to outdoor, outdoor to indoor, walk upstairs, walk downstairs), and (iv) *Postural/ Stationary* (just stand, lie on bed, sit on bed, lie on floor, sit on floor, lie on sofa, sit on sofa, sit on commode, lean on wall) |

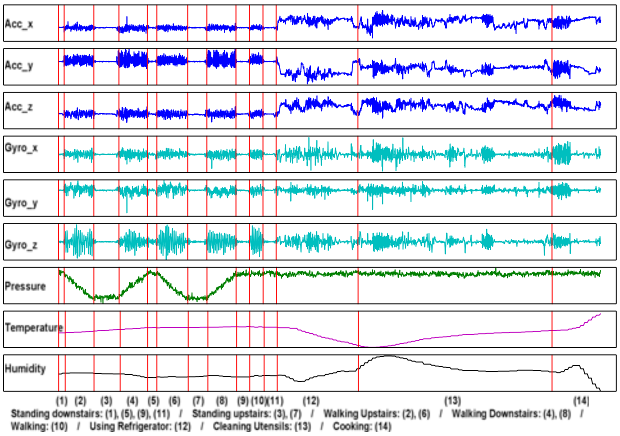


Fig. 2. Motivation behind utilizing multi-modal sensor data in the complex activity classifier, *HuMAn*. The figure shows variation of raw sensor data (accelerometer, gyroscope, air pressure, temperature, and humidity) from the smartphone wearable worn on wrist during sequence of different activities.

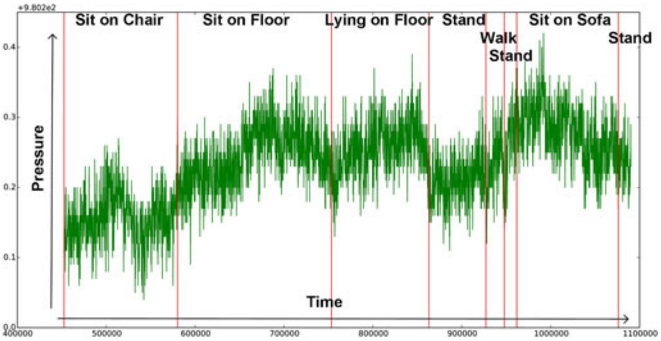


Fig. 3. Plot of barometric air pressure sensor data versus time, showing subtle variation of atmospheric pressure (from the smartphone wearable worn on thigh) during different activities inside a home environment on the ground floor.

## 3 HuMAn: A Complex Activity Recognition System

This section describes our complex activity classification system, *HuMAn*. Figs. 2 and 3 highlight the rationale behind this system. Fig. 2 demonstrates how a multi-sensor wearable worn just on the wrist has the potential to provide fine-grained signatures for activity detection ranging from stationary activities (e.g., sitting and standing) to activities like walking upstairs or downstairs, and more complex activities (e.g., cooking or cleaning utensils). The sensing information comes from multiple sensor modalities including accelerometer, gyroscope, atmospheric pressure, temperature and humidity sensors. In Fig. 3, we observe the subtle yet distinct signatures in variations of atmospheric pressure sensor data from a multi-sensor wearable worn on the thigh, for different complex indoor activities. In both cases, a smartphone is used as the data collection platform, but the same rationale holds true for any other multi-sensor wearable platform. If these subtle signatures can be combined with a) multi-modal information from other sensors placed on different body positions and b) location context information from miniature Bluetooth beacons, then it is possible to achieve significantly higher accuracy in detecting complex activities. This is the principal rationale and motivation behind our *HuMAn* system.

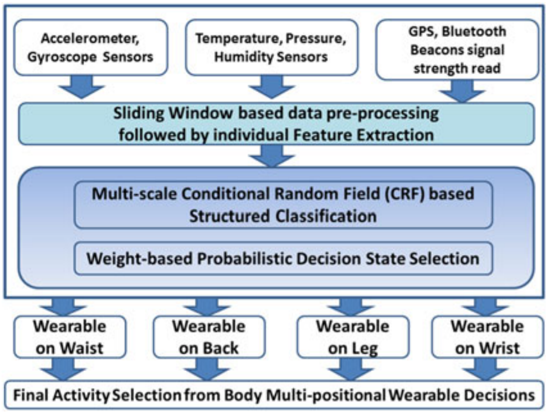

Accelerometer, Gyroscope Sensors
Temperature, Pressure, Humidity Sensors
GPS, Bluetooth Beacons signal strength read
Sliding Window based data pre-processing followed by individual Feature Extraction
Multi-scale Conditional Random Field (CRF) based Structured Classification
Weight-based Probabilistic Decision State Selection
Wearable on Waist
Wearable on Back
Wearable on Leg
Wearable on Wrist
Final Activity Selection from Body Multi-positional Wearable Decisions

Fig. 4. *HuMAn*: Complex activity classifier system.

### 3.1 System Overview

As illustrated in Fig. 4, the *HuMAn* system consists of three phases:

(1) Data pre-processing followed by feature extraction on each of the sensor data streams (executed separately on each wearable's data at multiple body positions).

(2) Multi-scale Conditional Random Field (CRF) classification followed by weight-based probabilistic decision state selection (executed separately on each wearable's data stream).

(3) Final user activity state classification by integrating individual decisions from each of the wearables.

All sensor data across different modalities are individually pre-processed and fed to the feature extraction algorithm. These extracted features are used as input to the multi-scale CRF classifier [31]. But instead of using a deterministic decision about activity states, our classifier employs a novel weight-based probabilistic activity state selection approach. This selection is done from the set of top $K$ classified activities and their emission/ output probabilities. Finally, the classifier decisions from the individual wearable devices at different body positions are integrated into a final activity state using a multi-positional selection approach. This last phase of decision selection in the *HuMAn* system is flexible, based on the number of wearable devices worn by the user on various body positions. This phase intelligently exploits the soft decisions from each device towards making an integrated final decision on the activity. Taking into consideration the complexity of integrating multi-modal sensor data from multiple body positions for decision making, our proposed system is designed for each device to make independent soft decisions that are then integrated for deciding on the final activity. This approach also leverages the processing power of each device. Adapting this design to recognize the final activity from directly fusing and processing multi-modal and multi-positional wearable sensor data is possible. However, there exist challenges such as packet losses, delays, jitter and time synchronization between devices, the exploration of which is part of our future work.

### 3.2 Two-Layer Classification Algorithm in *HuMAn*

The workflow of the *HuMAn* system is formally presented in Algorithm 1. As shown in *Step 1*, raw training data $TrD_i$ and testing data $TeD_i$ from multi-modal sensors on each wearable device $i$ (placed on four body positions: waist, lower back,

TABLE 3
Features Initially Calculated from All of the Available Raw Sensor Data

| Accelerometer Features | Components | | |
|---|---|---|---|
| | $\mu = Mean, \rho = Variance, \sigma = Std.Dev$ | | |
| Resultant | $\mu_{acc_R}$ | $\sigma_{acc_R}$ | |
| First Derivative | $\mu_{acc_{fd}}$ | $\sigma_{acc_{fd}}$ | |
| Second Derivative | $\mu_{acc_{sd}}$ | $\sigma_{acc_{sd}}$ | |
| Correlation | $\rho_{acc_{xy}}$ | $\rho_{acc_{yz}}$ | $\rho_{acc_{zx}}$ |
| Square Mean | $\mu_{acc_{x^2}}$ | $\mu_{acc_{y^2}}$ | $\mu_{acc_{z^2}}$ |
| Square Variance | $\rho_{acc_{x^2}}$ | $\rho_{acc_{y^2}}$ | $\rho_{acc_{z^2}}$ |
| Square Sum Mean | $\mu_{acc_{x^2+y^2}}$ | $\mu_{acc_{y^2+z^2}}$ | $\mu_{acc_{z^2+x^2}}$ |
| Square Sum Variance | $\rho_{acc_{x^2+y^2}}$ | $\rho_{acc_{y^2+z^2}}$ | $\rho_{acc_{z^2+x^2}}$ |
| **Gyroscope Features** | | | |
| Resultant | $\mu_{gyro_R}$ | $\sigma_{gyro_R}$ | |
| First Derivative | $\mu_{gyro_{fd}}$ | $\sigma_{gyro_{fd}}$ | |
| Second Derivative | $\mu_{gyro_{sd}}$ | $\sigma_{gyro_{sd}}$ | |
| Correlation | $\rho_{gyro_{xy}}$ | $\rho_{gyro_{yz}}$ | $\rho_{gyro_{zx}}$ |
| Square Mean | $\mu_{gyro_{x^2}}$ | $\mu_{gyro_{y^2}}$ | $\mu_{gyro_{z^2}}$ |
| Square Variance | $\rho_{gyro_{x^2}}$ | $\rho_{gyro_{y^2}}$ | $\rho_{gyro_{z^2}}$ |
| Square Sum Mean | $\mu_{gyro_{x^2+y^2}}$ | $\mu_{gyro_{y^2+z^2}}$ | $\mu_{gyro_{z^2+x^2}}$ |
| Square Sum Variance | $\rho_{gyro_{x^2+y^2}}$ | $\rho_{gyro_{y^2+z^2}}$ | $\rho_{gyro_{z^2+x^2}}$ |
| **Ambient Features** | | | |
| Pressure | $\mu_p$ | $\sigma_p$ | |
| Temperature | $\mu_t$ | $\sigma_t$ | |
| Humidity | $\mu_h$ | $\sigma_h$ | |
| **Bluetooth Beacon Features** | | | |
| Location index | $I_{loc}$ | | |

thigh, and wrist) are passed through low-pass and median filters for noise reduction and smoothing. Then critical features $F_i$ are extracted from processed data as described in Section 4.

In *Step 2*, the pre-processed training data features paired with ground truth activities $A_i$ from all 4 wearables denoted as $\left[F_i, A_i\right]$ for $1 \leq i \leq 4$, are used for training the CRF model. Subsequently, the weight $W_{ik}$ is calculated for each device-activity pair. Higher $W_{ik}$ implies that device $i$ has higher confidence to predict the activity $k$. The CRF parameter $\omega$ is estimated using maximum likelihood function to maximize the conditional probability $P\left(A_i|F_i;\omega\right)$. The execution of CRF is described in Section 5.2.

In *Step 3*, features $\left[F_i\right]$ for $1 \leq i \leq 4$, extracted from the testing data are fed to the CRF model to calculate the conditional probability $P\left(A_i|F_i;\omega\right)$. The conditional probability of each activity is multiplied with weight $W_{ik}$ to calculate the final activity score. The activity having the highest score is selected as inference from the corresponding device. Finally, the multi-positional decision selection approach (see Section 6) is applied to select the final decision from four soft decisions from each smartphone.

## 4 Sensor Sampling and Feature Selection

This section discusses in detail the sensor sampling, feature extraction and feature selection process used in *HuMAn*. This is described in Algorithm 1.

### 4.1 Sensor Sampling

*Accelerometer and Gyroscope*. In our system, the 3-axis accelerometer and 3-axis gyroscope are both sampled at 100 Hz. This sampling frequency is enough to capture human body movements [19]. The *HuMAn* system is designed to compute features from these two modalities for each sampled sliding window of size two seconds.

TABLE 4
Filtered Features by Applying Relief-F and Correlation-Based Evaluation Algorithm

| Feature selection method | Selected Features |
|---|---|
| Relief-F | $\mu_{acc_{z^2+x^2}}$, $\mu_{acc_{z^2}}$, $\mu_{acc_{y^2}}$, $I_{loc}$, $\mu_{acc_{x^2+y^2}}$, $\mu_t$, $\mu_h$, $\mu_p$, $\mu_{acc_{x^2}}$, $\mu_{gyro_R}$, $\rho_{acc_{zx}}$, $\mu_{acc_{y^2+z^2}}$, $\rho_{acc_{yz}}$ |
| Correlation-based Evaluation | $\mu_{acc_{sd}}$, $\sigma_{acc_{sd}}$, $I_{loc}$, $\mu_t$, $\mu_h$, $\mu_p$, $\mu_{acc_{x^2}}$, $\mu_{acc_{y^2}}$, $\mu_{acc_{z^2}}$, $\mu_{acc_{y^2+z^2}}$, $\mu_{acc_{z^2+x^2}}$, $\mu_{gyro_{x^2+y^2}}$, $\rho_{acc_{yz}}$, $\rho_{acc_{zx}}$, $\rho_{gyro_{xy}}$, $\rho_{gyro_{yz}}$, $\rho_{gyro_{zx}}$ |

*Temperature, Humidity and Air Pressure sensors*. These ambient sensors are sampled at 1 Hz, 1 Hz and 5 Hz, respectively. Since they are sampled at low frequencies, they do not add significant burden on energy consumption but add valuable relevant information for activity recognition. The pressure sensor sampling rate is set a bit higher to capture fine changes in atmospheric pressure in different location contexts as shown in Figs. 2 and 3.

*Location Context*. The *HuMAn* system is targeted for complex activity recognition in indoor homes. It is designed to use GPS sensor (if available) and Bluetooth message receptions from beacons deployed in the infrastructure which are sampled at 1 Hz frequency. Although GPS signals are usually not available, or are incorrect indoors, *HuMAn* collects GPS data to help recognize activities like outdoor to indoor transitions. But the more effective location features are Bluetooth beacon message containing RSSI values. This is one of the novelties of our system. The simple, small and cheap Bluetooth beacon devices [18] are popular in commercial sectors. They can periodically notify nearby devices of their presence, thus representing proximity of those devices to the beacons. We have exploited this in *HuMAn* to enable location context based features. From the beacon id and corresponding RSSI values, it is feasible for nearby wearable devices to infer coarse-grain location contexts such as bedroom, kitchen, bathroom, etc.

### 4.2 Feature Selection

Feature selection is a very important part of a machine learning algorithm. It defines a mapping function between input features and output class based on the information from input features. Unfortunately, not every input feature provides useful information about the output class. Irrelevant features can cause problems like over-fitting, overhead, and inability to visualize the feature map to glean insights on the data.

In this paper, we have initially identified a pool of 49 features (Table 3) which are easy to compute in real-time. We have used filter-based approach using Relief-F [29], Wrapper-based approach using a Sequential Forward Floating Search (SFFS) [23], and Greedy approach based on Pearson's correlation co-efficient [20] to find a subset of relevant features which is a good balance of generality and performance. We also used many classifiers to evaluate error estimation, for example, K-Nearest Neighbors [38], Naive Bayesian [36], C 4.5 [39] and Random Forest [30]. These results are illustrated in Table 4 and Table 5. Finally, as shown in Table 6, a total of 12 features were selected based on the evaluations of these algorithms. All input features were normalized to obtain best results for the classifier used

TABLE 5
Filtered Features by Applying Wrapper Based Algorithm

| Wrapper-based Feature selection method | |
|---|---|
| Classifier | Features |
| Naive Bayesian | $\mu_{gyro_R}$, $I_{loc}$, $\mu_t$, $\mu_p$, $\rho_{acc_{yz}}$, $\rho_{acc_{zx}}$, $\rho_{gyro_{x^2}}$, $\mu_{acc_{y^2+z^2}}$, $\mu_{acc_{z^2+x^2}}$ |
| K Nearest Neighbors (KNN) | $\mu_t$, $\mu_h$, $I_{loc}$, $\mu_{acc_{x^2}}$, $\mu_{acc_{y^2+z^2}}$ |
| Random Forest | $\mu_{gyro_R}$, $\mu_t$, $\sigma_t$, $\mu_h$, $I_{loc}$, $\rho_{acc_{zx}}$, $\mu_{acc_{x^2}}$, $\sigma_{acc_{x^2+y^2}}$, $\mu_{acc_{z^2+x^2}}$ |
| C 4.5 | $\mu_{acc_{sd}}$, $\mu_t$, $I_{loc}$, $\mu_{acc_{z^2+x^2}}$ |

TABLE 6
Finally Selected Features, Based on the Ones Common in All or at Least in Multiple Classifiers

| Final Selection of Features | |
|---|---|
| Classifier | Features |
| Almost common in all classifiers | $I_{loc}$, $\mu_t$, $\mu_p$, $\mu_{acc_{z^2+x^2}}$, $\mu_{acc_{y^2+z^2}}$, $\rho_{acc_{zx}}$ |
| Common in multiple classifiers | $\mu_{acc_{x^2}}$, $\mu_{acc_{y^2}}$, $\mu_{gyro_R}$, $\rho_{acc_{zx}}$, $\mu_{acc_{y^2}}$, $\mu_{acc_{z^2}}$, |

for feature evaluation. This is to ensure equal weight to all the potential features, thus reducing bias. The feature selection techniques are briefly presented below.

1) *Filter-based feature selection using Relief-F*: This algorithm relies on contextual information and dependencies between the features to estimate the quality of features. Initially, it assigns a default weight to each feature. Then, while iterating through each data point, it increases the weight of those features which exhibit significant difference in values for different classes, and then decreases the weight for those that are unchanged for different classes. The process is repeated for $p$ times, where $p$ is a user-defined parameter, and finally it calculates the average weight for each feature. A higher weight for a feature means more utility for classification [29].
2) *Wrapper-based feature selection using SFFS*: The basic difference between the filtering and wrapper methods is that the filtering method evaluates subsets by their information content (e.g., interclass distance, nearest neighbor, statistical dependence), whereas the wrapper-based method uses a classifier to evaluate subsets by their predictive accuracy (on test data) by statistical re-sampling or cross-validation. One of the disadvantages of a wrapper-based method is its lack of generality when applied to multiple classifiers [23].
3) *Correlation-based feature selection using greedy method*: Correlation based methods evaluate the worth of a subset of features by considering the predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation between other features are preferred. The method uses Pearson's correlation coefficient [6] to evaluate the subset correlations. Finally, a greedy forward or backward search is made through the space of feature subsets. It starts with no/all features or from an arbitrary point in the space, and stops when the addition/deletion of any remaining features results in a decrease in the final evaluation [21].

## 5 Structured Classification

We now describe the graphical model based structured classifier used in *HuMAn*, as presented in *Step 2* of Algorithm 1.

### 5.1 Context-Based Classifier Selection

Selecting a classifier which suits the overall context well is very important. In this paper, we leverage the fact that human activities are generally sequential in nature and exhibit spatial-temporal properties. An activity performed at any given time instance is highly influenced by those performed in previous time instances. For example, a person walking at any moment is most likely to continue to walk in the very next second as well. By considering this context carefully we selected (and adapted) the idea of Condition Random Fields (CRF) algorithm for our problem scope, because unlike other supervised learning algorithms, the CRF model makes predictions based on not only the current observation but also on past observations and future predictions. Details of the CRF algorithm are described next.

### 5.2 Conditional Random Fields

CRFs are a class of statistical modeling methods used for structured learning and prediction [31]. It is a discriminative counterpart model for generative Hidden Markov Model (HMM) algorithm. While HMM leverages the sequential nature of data, the CRF does the same but with more general assumptions compared to HMM. Moreover, HMM defines dependency between each state and "only" corresponding observations, whereas CRF models the dependence between each state and the entire observation sequence. The *HuMAn* system utilizes this feature by developing a structured classification approach for complex activity recognition. In our version, the notion of CRFs is leveraged via an undirected graphical model based design in order to label sequences of fine-grained activity data. It allows seamless integration of varied features from multimodal sensor data into the graphical model.

The CRF model is formally defined as follows. Let $\vec{x} = (x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2})$ be a sequence of input feature vectors, where $x_t$ represents the feature vector extracted from raw multi-modal sensor data at time $t$. Let $\vec{y} = (y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2})$ represent the corresponding sequence of activities performed. Let $L$ be the length of the sequence. The goal of CRF is to learn a good mapping from $\vec{x}$ to $\vec{y}$ given a training set of N samples. To do so, the CRF computes the conditional probability

$$P(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{x})} exp \sum_{j=1}^{L} \omega \cdot \Phi_j(\vec{y_t}, \vec{y_{t-1}}|\vec{x}). \tag{1}$$

Here, $\omega$ is a weight vector trained by the training data with the objective to maximize $P(\vec{y}|\vec{x})$, and $\Phi_j(\vec{y_t}, \vec{y_{t-1}}|\vec{x})$ is a feature mapping function which defines the dependency between input and output vectors. The graphical representation of feature mapping function is described in the next section. $Z(\vec{x})$ is a partition function which acts as a global normalizer so that Eq. (1) yields a valid probability. It is given by,

$$Z(\vec{x}) = \sum_{y} \left( exp \sum_{j=1}^{L} \omega \cdot \Phi_j(\vec{y_t}, \vec{y_{t-1}}|\vec{x}) \right). \tag{2}$$

**Algorithm 1.** CRF-Based Algorithm for *HuMAn* System

$TrD_i$ = Training data from sensors on $i$-$th$ device
$TeD_i$ = Testing data from sensors on $i$-$th$ device
$F_i$ = Features extracted from $TrD_i$ on $i$-$th$ device
$A_i$ = Classified Activity from $i$-$th$ device
$P(A_i|F_i)$ = Conditional Probability of Activity $A_i$ given feature $F_i$

**Step 1 Pre-Processing**:

1) Median filters are applied to remove accidental spikes from $TrD_i$ and $TeD_i$.
2) Low-pass filters are applied to remove high frequency signals from $TrD_i$ and $TeD_i$.
3) Features $F_i$ are extracted from processed data $TrD_{ip}$ and $TeD_{ip}$ obtained from steps (1) and (2).

**Step 2 Training**:

**Input**: Training dataset tagged with ground truth $\left[TrD_i, A_i\right]_{i=1}^{i=4}$

**Output**: Trained CRF parameters $\omega$ from Section (5.2) Eq. (1) and activity weight $\left[W_{ik}\right]_{i=1,k=1}^{i=4,k=21}$ for each activity $\left[A_{ik}\right]_{i=1,k=1}^{i=4,k=21}$ for every device, where $i$ and $k$ represent device id and activity id, respectively.

1) Manually decide on the feature functions $\Phi_j(A_t, A_{t-1}|F)$, i.e., connections between different activity windows $A_t$ and $A_{t-1}$ at time $t$ and $t-1$ respectively, given input features $F$ as shown in Fig. 5.
2) Given all training pairs $\left[TrD_{ik}, A_{ik}\right]_{i=1,k=1}^{i=4,k=21}$, apply conditional maximum likelihood to find the optimal $\omega$ to maximize $P\big(A_i|F_i;\omega\big)$.
3) Estimate the weight parameter $W_{ik}$ based on the precision of each device $i$ for corresponding activity $k$.

**Step 3 Prediction**:

**Input**: Testing dataset without tagged ground truth $\left[TeD_i\right]_{i=1}^{i=4}$, Trained CRF parameter $\omega$, Feature functions $\Phi_j(A_t, A_{t-1}|F)$ and Activity weight parameter $W_{ik}$.

**Output**: Final activity selection $A_{fs}$

1) Calculate conditional probability of each activity given input features using Eq (1), $P\big(A_i|TeD_i;\omega\big)$.
2) Select the $K$ top activities based on conditional probability values.
3) Select the final activity from each device $A_{ik} = (W_{ik} * \frac{P_{ik}}{\sum_k P_{ik}}) > 0.5\ \forall i \in [1,4]$ and $\forall k \in [1,21]$.
4) Select final activity $A_{fs}$ from one of $i$ decisions, one from each device, based on the physiological context evaluation from each device-activity pair.

### 5.3 Graphical Structure

The CRF model captures the temporal relationships in sequential activity data. The graph structure used in the CRF model is illustrated in Fig. 5. It shows the observation sequence $\vec{x}$ (obtained from the multi-modal sensor feature extraction phase in the *HuMAn* system), hidden states $\vec{y}$ (activity states) of class probability assignments, and the edges $E$ between hidden states representing pairwise relationships. As in Fig. 5, the different scales or lengths of edges (e.g., that run from $y_t$ to $y_{t-1}$, $y_{t+1}$, $y_{t-3}$, $y_{t+3}$, $y_{t-4}$, $y_{t+4}$) enable the flow of contextual information in the whole network. When we say the feature function template is "010304", we mean that in order to detect the activity in current time window, the model not only relies on the current input time window but also on the output and input of the immediately prior and next (in this case, 1st, 3rd and 4th) data windows. This relationship template enables the model to be more informed and considers the temporal relationship between the activities more effectively. For the implementation of CRF model, we have used and modified the standard CRFSharp toolkit [14].

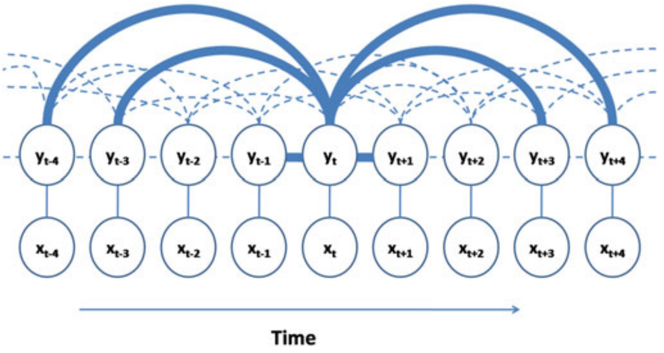


Fig. 5. CRF graphical structure. The thick edges represent the pairwise edges for the template setting of $(010304)$ for hidden state node $y_t$. We use a slightly different CRF template naming convention than the one in [53].

## 6 Multi-Positional Decision Selection

This process is indicated in *Point 4* under *Step 3* of Algorithm 1. Note that each of the wearable devices in our *HuMAn* system will independently classify the activity performed. While in most cases, we expect coherence in the activities classified, in some cases, the same activity may not be classified by all of the wearables. When this happens, the confusion needs to be resolved. We propose a simple approach to fix this issue based on fundamental insights on human activity (that were gleaned from discussion with experts in human kinesiology). In our approach, we assign a simple relevance index for each activity against the position of the wearable. For example, cleaning utensils has relevance to the wearable on the wrist and thigh; lay on sofa has relevance to the wearable on the waist, back and the thigh; while standing has relevance to the wearable on the thigh. In our system, when discrepancies happen, the final activity $A_{fs}$ is chosen from the set $\bigcup_{k\in\{1\ldots21\}}\{(A_k, i)\}$ by considering only the pairs $(A_k, i)$ where activity $A_k$ is relevant to $i$th wearable's position on the body as assigned previously. In the rare case of ties, $A_{ik}$ is chosen randomly from the equally probable choices.[1]

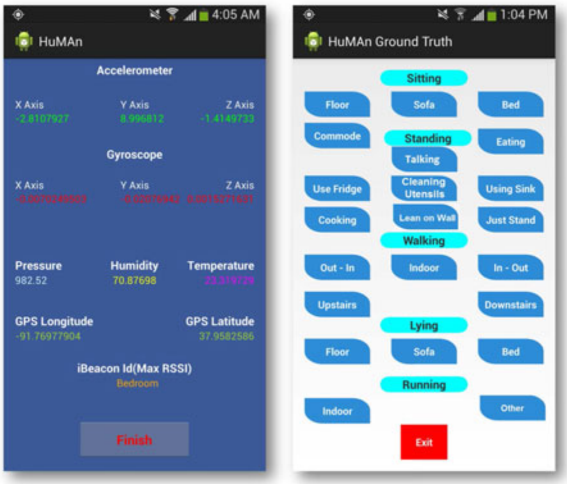
## 7 Experimental Evaluation and Analysis

In this section we present the experimental evaluation and performance validation of the *HuMAn* system.

### 7.1 Experiment Setup

We have tested activity recognition performance of *HuMAn* with a total of 40 datasets from 10 adults (3 female and

[1] Considering the complexity of our problem scope from the number of complex activities, sensor modalities, and their emplacements in various positions of the body, we opted for this approach to resolve discrepancies. Investigating this issue more thoroughly from the kinesiology perspective is part of our future work.



4:05 AM

HuMAn

Accelerometer

X Axis -2.8107927
Y Axis 8.996812
Z Axis -1.4149733

Gyroscope

X Axis -0.0070249503
Y Axis -0.02076942
Z Axis 0.0015271631

Pressure 982.52
Humidity 70.87698
Temperature 23.319729

GPS Longitude -91.76977904
GPS Latitude 37.9582586

iBeacon Id(Max RSSI) Bedroom

Fig. 6. Smartphone app *HuMAn* (on left) developed for multi-sensor data collection. Smartphone app *HuMAn Ground Truth* (on right) developed for ground truth data logging by an external observer.

7 male) aged between 20 and 25. To minimize location biases, the experimental data for 4 subjects were collected from one home, and for the remaining 6 subjects from another home. Both experimental locations are duplex apartments with bedrooms, kitchen, washrooms, stairs, etc. There are four datasets collected concurrently from each user as they perform activities. That is, one dataset from each of the four smartphones worn at the waist, back, wrist and thigh. Each phone senses data from all three categories of sensors: activity, ambiance and location. For the *HuMAn* system deployment, we have used Samsung Galaxy S4 [44] smartphones and onboard sensors (for proof of concept), along with Bluetooth beacons from Gimbal Inc. [18] deployed in the external infrastructure. Note that smartphones are used only as a "minimum viable product (MVP)" for multi-sensor data collection platform in this paper. Our algorithms are agnostic to the actual wearable device placed on the body.

We developed an Android application for the *HuMAn* system which senses data from selected onboard sensors in the phone, and receives Bluetooth signals from beacons installed in different rooms of the home. The data are locally stored on the smartphone with proper timestamps. We developed another Android application *HuMAn Ground Truth* for collecting ground truth with proper timestamps. Screenshots of these apps are shown in Fig. 6. The *HuMAn* application was installed on the 4 phones worn by the subject, while the *Human Ground Truth* application was installed on an external observer's smartphone for recording the ground truth. To do so, the observer taps the buttons corresponding to every activity to record the start time and end time for each of those activities. Both applications use time synchronization from the Network Time Protocol (NTP) [35] server for finer accuracy with timestamps. All the users were instructed to naturally perform the set of activities in any order and duration of their choice. The timestamps of activities as logged by the observer, and the timestamped sensor logs in the 4 smartphones were merged to prepare the training dataset.

Note that, while the observer tagged the 21 activities correctly as they are performed by a subject, each subject also did one or more 'other' activities between the end time and the start time of certain consecutive activities among the 21 of interest. This is natural, as it represents uniqueness of each individual transitioning between the 21 activities. The 'other' activities were diverse across users (male and female). However, the most common ones were 'adjusting hair', 'opening and closing kitchen cabinets', 'adjusting spectacles', 'washing hands in faucet', and 'drinking water'. The observer carefully tagged all such activities outside the 21 of interest as 'other', as they were being performed by subjects in our experiments. Our classifier is trained to classify such activities simply as 'other' for overall completeness.

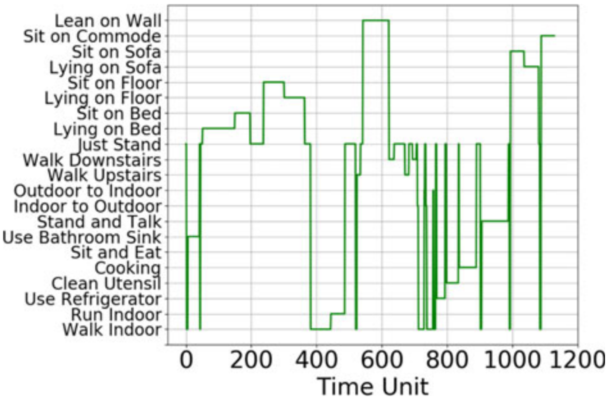


Fig. 7. Ground truth activity sequence performed by User-1.

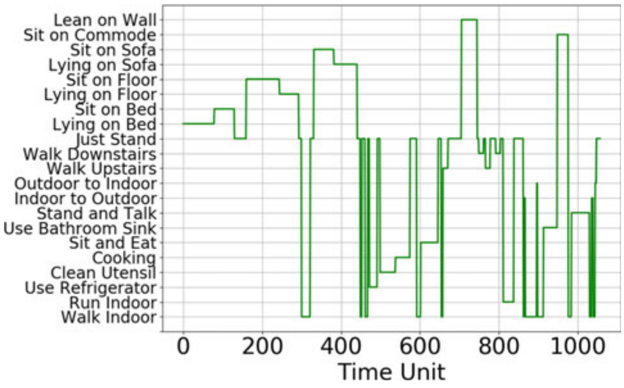


Fig. 8. Ground truth activity sequence performed by User-2.

For illustration, only the sequence of those 21 activities performed by two arbitrary users are presented in Figs. 7 and 8. Each user's duration to perform all activities took an average of 45 minutes. To summarize, from all users and smartphones, we collected a total of about 28 million data points from various sensors in our dataset.

### 7.2 Classifier Design

Let us now describe the parameter and configuration setup of the activity classifier proposed in *HuMAn*. For feature extraction from raw sensor data, we have used the sliding window approach where the window size is varied from 1 second to 10 seconds with different amounts of overlap. After observations, we chose a sliding window size of 2 seconds with 50 percent overlap. For every sliding window, the system calculates the best 12 features from raw sensor data of 7 sensors, as discussed in Section 4.

We have also optimized the configuration in the CRF classifier graph structure. CRF allows to build a relationship function from current sample data point to the previous and next data points in time to produce contextually better classification models. This procedure is mostly problem

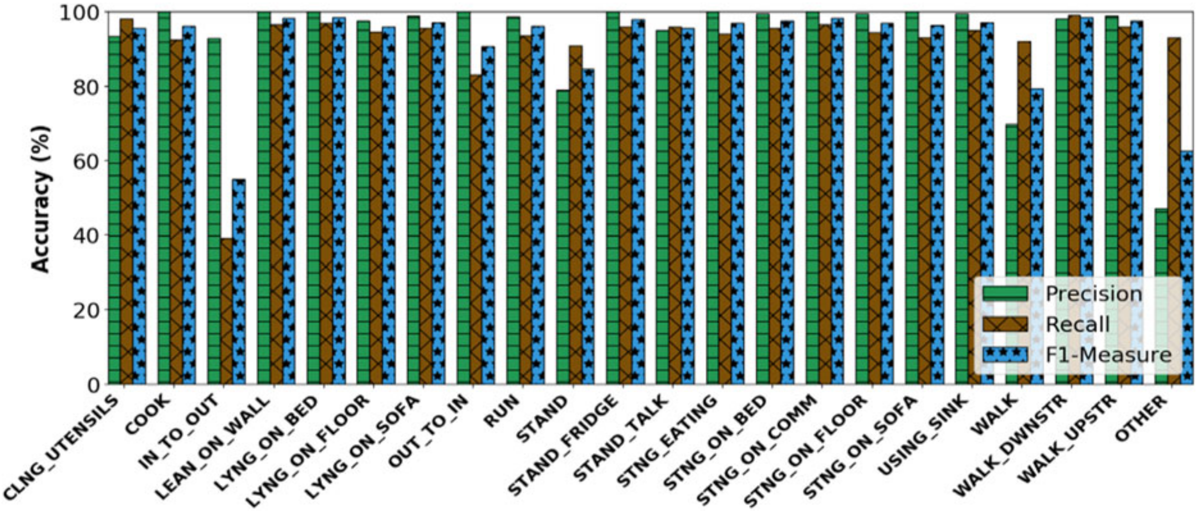


Fig. 9. Same-User 10-fold cross-validation performance evaluation for each activity from multi-positional data.

specific and is standard in CRF based designs. We have evaluated *HuMAn* with different edge configurations of the graph and found out that the best configuration is for the case where the feature function is generated by considering the 1st, 4th, 9th and 19th data samples – both preceding and succeeding in the graph structure, which is used throughout in this paper. In our design, during training, we use our dataset to find the precision of each wearable (placed on different body positions) to correctly detect an activity. During classification, instead of selecting the best activity based on CRF execution from each wearable, we select the corresponding top 3 activities, and multiply their probabilities with the precision derived during training to choose one final activity from each wearable. Finally, the activity with a probability greater than 0.5 gets labeled as the final one for corresponding window from the respective wearable. In case none of the activities have a probability greater than 0.5, the final activity was labeled as 'other' appropriately. In this way, the system is a little more complete, since any activity outside of the 21 of interest will simply be classified as 'other' as discussed earlier in Section 7.1. Unless otherwise stated, the classification results are presented only for the implementation of the multi-positional approach in Section 6 by considering results from the four wearables in all of the four body positions. In the next three sections, three different evaluation strategies and their results are presented for *HuMAn* system. For each evaluation strategy, Precision, Recall and F1-measure scores have been plotted for each activity. Low precision for any activity means the classifier predicts too many false positive results for that particular activity, whereas low recall for any activity means that the classifier predicts too many false negative results for that activity. The F1-measure combines both of them, and calculates the harmonic mean of precision and recall.

### 7.3 Same-User 10-Fold Cross-Validation Evaluation

In this evaluation strategy, the dataset from each subject is independently trained and tested. Here, the dataset belonging to each subject is split into 10 sections of data randomly. Each of the 9 sections of data is used to train the model and the remaining one section is used for evaluation. Finally, the results from all 10 sections were averaged to evaluate the final Precision, Recall and F1-Measure of classification for each subject. We also show the confusion matrix [45], which indicates the error distribution in classification across 21 activities. For each matrix element, the corresponding row indicates the actual class (i.e., the ground truth activity) while the corresponding column indicates the predicted class or activity with the classification accuracy in percentage (indicated by the value of the matrix element). Higher values along the diagonal entries indicate better accuracy.

Results and confusion matrix for same-user evaluation are shown in Figs. 9 and 10. As can be observed, most activities are correctly classified. Only the "Outdoor to Indoor" and "Indoor to Outdoor" transition activities are less accurate and model predictions are confused between these two. This is because for these two activities the model heavily relies on information from the environmental sensor, since the ambiance between outdoors and indoors are different. But ambient sensors like the temperature sensor have more sensing delays, which adds noise, thus lowering the accuracy. Additionally, we observe that in selected cases, there is minor confusion in closely related activities, such as Walking Upstairs versus Standing, and Sitting on Floor versus lay on Floor, which is understandable since these activities are very closely related.

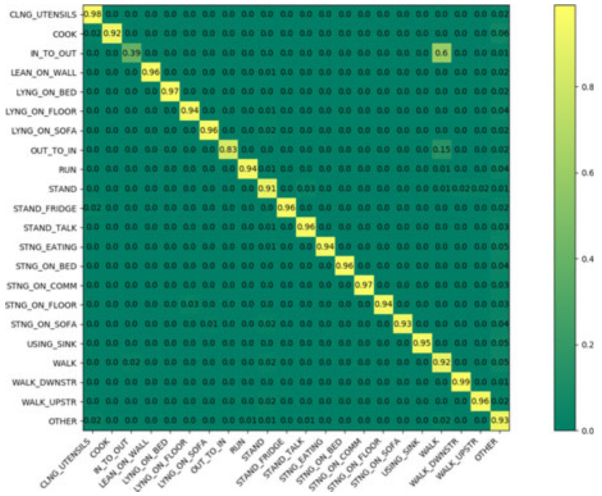


Fig. 10. Confusion matrix for same-user 10-fold cross-validation evaluation.

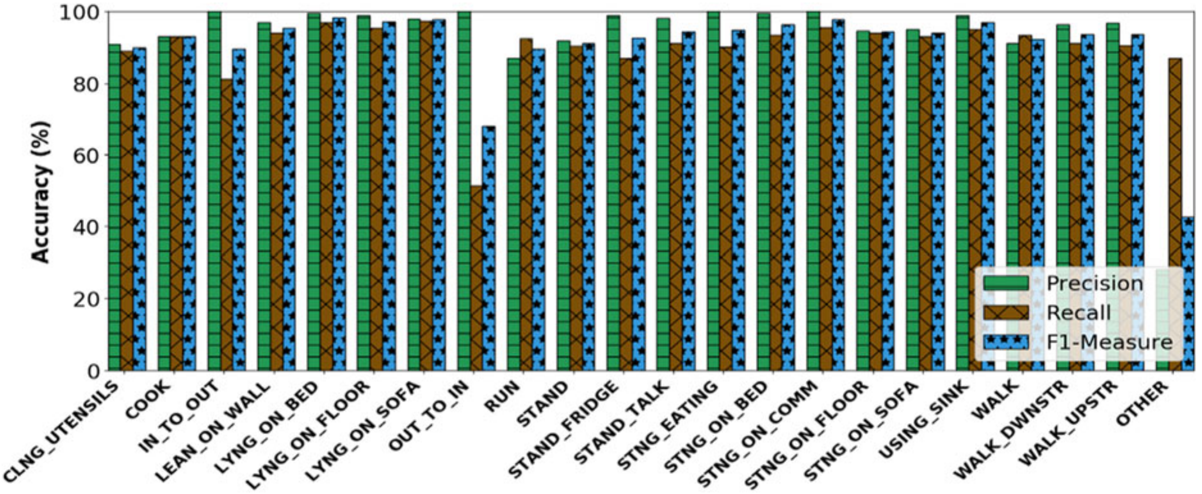


Fig. 11. Cross-user 10-fold cross-validation performance evaluation from multi-positional data.

To summarize, for the same-user evaluation strategy, the overall accuracy is 95.38 percent, while the accuracy ranges from 60.25 to 98.90 percent across various activities. But 17 of those activities have accuracy higher than 90 percent and the median of accuracies across all activities is 93.17 percent.

### 7.4 Cross-User 10-Fold Cross-Validation Evaluation

This is a stricter evaluation technique. Here, we combine data from all users in a common dataset pool to apply cross-validation (instead of separating datasets of each user). The combined data is divided into 10 sections where each was tested against the model of remaining 9 sections of training data. Typically, such datasets have more noise since even the same activity will be performed a little differently by different users, which is more so for complex activities. Fig. 11 shows that the overall activity classification accuracy for this evaluation strategy is 92.29 percent. The error distribution can be seen in the confusion matrix in Fig. 12. From this figure, it can be seen that for this evaluation strategy also, our model performs very accurately and most confusion happens in such activities as “Indoor to Outdoor” and “Outdoor to Indoor”. For other activities, there is a degree of confusion, wherein many motion activities are confused with the “Standing” activity.

### 7.5 Cross-User Leave-One-Out Cross-Validation Evaluation

Cross-user leave-one-out cross-validation is the strictest evaluation strategy. Here, we separate the data of subjects used for training from those for testing. Therefore, completely new unseen data are tested against the model built with orthogonal training data. The accuracy and Confusion matrix are shown in Figs. 13 and 14, respectively. The overall accuracy here is 74.49 percent. Specifically, four activities are less than 60 percent accurate while 13 activities have more than 80 percent accuracy.

We can observe from Fig. 13 that activities on the right side of the figure have higher accuracies in classification compared to the ones on the left. This is because activities on the right hand side are mostly atomic in nature and hence can not be broken into simpler activities. For such activities, the way they are performed are similar across users, which leads to less inter-subject variability and higher detection accuracy. On the other hand, activities on the left of Fig. 13 are more complex and composed of multiple atomic activities, the presence of which lead to gradually more inter-subject variabilities as they are being performed. This is especially true for activities like cleaning utensils, cooking, lay down (on floor versus sofa versus bed) etc., leading to lower classification accuracies. Encouragingly though, from the corresponding Confusion Matrix in Fig. 14, we observe that these complex activities are confused with *similar* atomic activities. For example, ‘cooking’ confuses most with ‘standing’; ‘cleaning utensils’ confuses the most with ‘standing’; and ‘sitting on sofa’ confuses with ‘lay on sofa’ etc. These trends give us confidence that our system is not too far off in classifying complex activities accurately. With more subjects, and more diversity across them, we expect significant performance improvement, which is part of our future work.

### 7.6 Impact of Sensing Modality and Sensor Placement

We now show the impact of integrating multiple sensor modalities, as well as the impact of sensor placements on

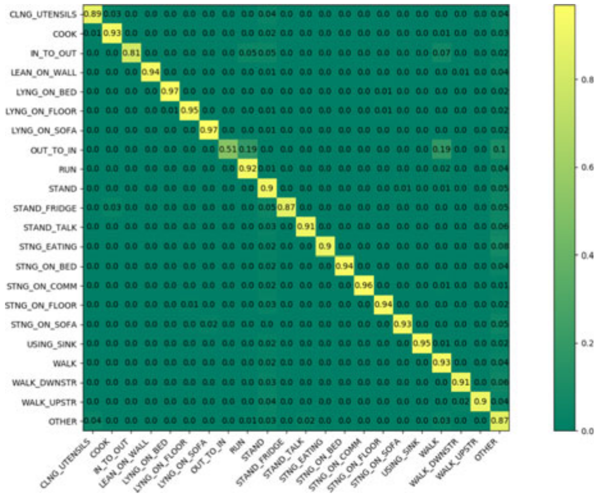


Fig. 12. Confusion matrix for cross-user 10-fold cross-validation evaluation.

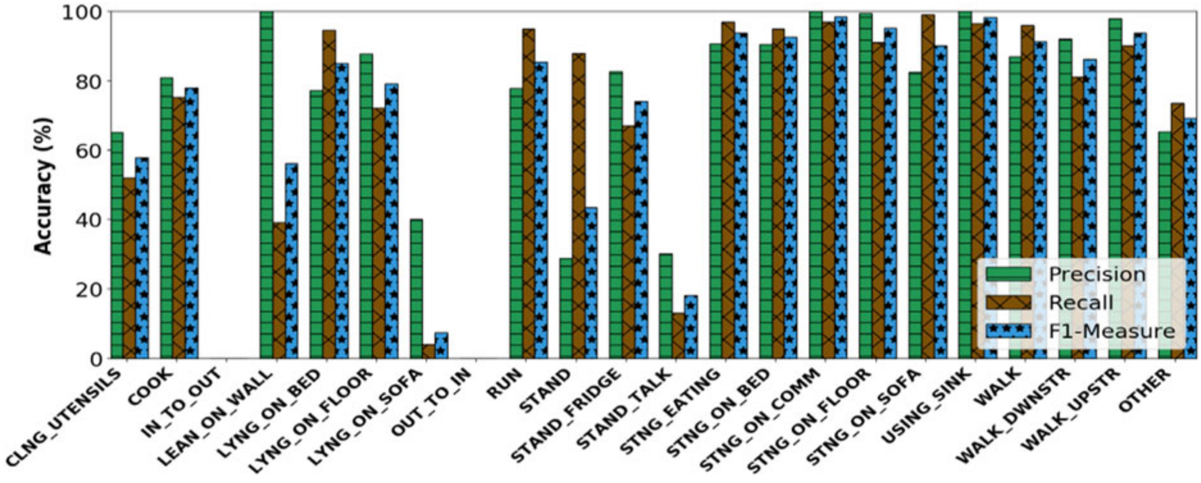


Fig. 13. Cross-user leave-one-out cross-validation performance evaluation from multi-positional data.

multiple body positions. Fig. 15 demonstrates the performance by evaluating classification accuracy with: (i) features from only movement activity sensing (i.e., accelerometer and gyroscope) denoted as “Activity”; (ii) features from movement sensing and ambient sensing modes (i.e., addition of temperature, air pressure and humidity sensors), denoted as “Activity + Ambiance”; (iii) features from movement sensing and location based sensing modes (i.e., from Bluetooth beacons in proximity), denoted as “Activity + Location”; and (iv) integrating features of above three contexts. From Fig. 15, we can observe that while adding features with more contextual information improves accuracy, the best classification accuracies of 95 percent for same-user and 75 percent for cross-user evaluation were achieved when features from all three contexts are integrated, hence validating our *HuMAn* system.

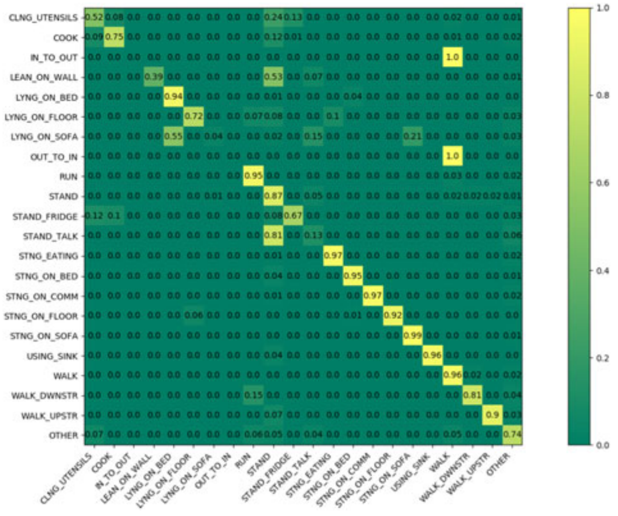


Fig. 14. Confusion matrix for cross-user leave-one-out cross-validation evaluation.

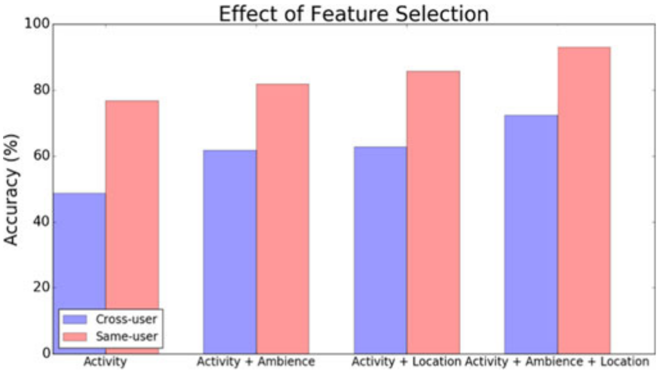


Fig. 15. Average accuracy performance of different combination of sensing modalities in *HuMAn* both for same-user 10-fold cross-validation evaluation, and cross-user leave-one-out cross-validation evaluation.

Finally, Fig. 16 presents the impact of integrating information from sensors placed in multiple body positions for classification. Here again, the highest accuracies in classification were obtained when information is integrated from sensors placed in multiple positions on the body, as opposed to the classification accuracy obtained by leveraging sensor streams from any single position on the body.

## 7.7 Sensor Energy Consumption and Algorithmic Time

The Samsung Galaxy S4 phone used in our experiments is equipped with a 2600 mAh battery rated at 3.8 V which is equivalent to 9.88 Wh. In our experiments, we see that for recording accelerometer and gyroscope readings at 100 Hz for about 45 minutes of continuous activity, the average energy expended was 6 ± 1 percent of the total energy in that phone. The energy consumption is hence manageable.

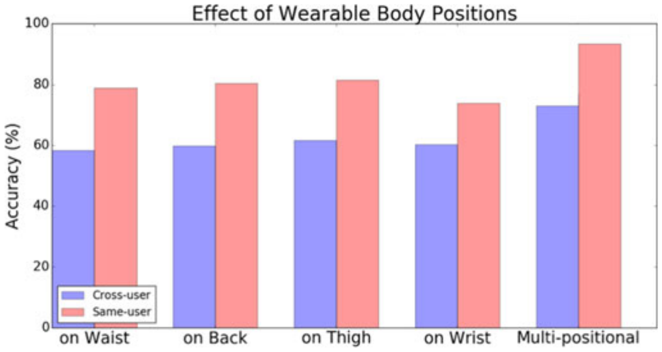


Fig. 16. Average accuracy performance of multi-positional decision making in *HuMAn*, compared to individual activity decisions from each device placed on different body positions for same-user 10-fold cross-validation evaluation, and cross-user leave-one-out cross-validation evaluation. Each device runs the *HuMAn* activity classifier.

We also present results on the time complexity of our CRF algorithm. Note that the time complexity of standard training for CRF is quadratic in the size of the output class, linear in the number of features, and quadratic in the size of the training sample. Similarly, the time complexity of inference for CRF is quadratic in the size of the output class. We performed all training and inference of the CRF algorithm on a server with Intel CPU i7-5600U and 2.60 Hz frequency. Typically it took a little less than 12 minutes to train the model with 10 users data, and 110 ms on an average for inferring a single input data unit. Naturally, the total inferring time in the CRF is proportional to the total number of classes in the model. Since the size of the class is already fixed, inferring takes constant time for each input unit. Implementing the CRF classification algorithm as a smartphone app is part of our ongoing efforts.

## 8 Conclusions and Discussions

In this paper, we designed *HuMAn*, a hybrid multi-modal sensor based and body multi-positional wearable context based complex activity recognition system. To the best of our knowledge, this paper is the first to design a system that can classify 21 fine-grained complex at-home activities with high accuracy. We leverage three different sensing contexts for multi-modal sensing: body locomotion, ambient environment, and location context. We exploit contextual information from sensors in multiple body positions to further improve activity classification. Experimental results demonstrate that for same-user evaluation strategy, the average activity classification accuracy is as high as 95 percent. For the case of 10-fold cross-validation evaluation strategy, the average classification accuracy is 92 percent, and for the case of leave-one-out cross-validation strategy, the average classification accuracy is 75 percent. We are currently investigating if addition of more data with more experiments across diverse users will help improve classification accuracies. Investigating gender-specific and age-specific models to extract novel features, and demonstrating improved classification accuracy is also part of our ongoing work.

Currently, we are enhancing *HuMAn* to classify more activities, and enabling the entire system to execute as a smartphone app with superior energy efficiency. We are also looking to better understand the healthcare impacts of our work from the perspective of activity recognition. To do so, we are actively discussing with healthcare professionals in the domain of dementia, cardiac care, and exercise therapy. Additionally, we are attempting real-time integration of multi-modal sensor data from multiple wearables and Bluetooth beacons at the same time to enhance classification accuracies. The challenge is how to handle wireless packet losses, compensating for network delays, jitter etc.

## Acknowledgments

The authors thank the anonymous reviewers and the Associate Editor for their constructive feedback and valuable suggestions to help improve the quality of the manuscript. This work was supported in part by the US National Science Foundation (NSF) under Award Nos. IIS-1404673, IIP-1648907, IIP-1540119, IIS-1254117 and II-New-1205695. D. De was a postdoctoral research fellow in computer science at the Missouri University of Science and Technology while this work was done.

## References

[1] *Biostrap Wearable Device*, [Online]. Available: https://www.biostrap.com/
[2] *Fitbit Wearable Device*. [Online]. Available: https://www.fitbit.com/
[3] *Lumo Back Wearable Device*. [Online]. Available: http://www.lumobodytech.com/lumoback/
[4] *Lumo Lift Wearable Device*. [Online]. Available: http://www.lumobodytech.com/lumolift/
[5] *Nike+ Wearable*. [Online]. Available: https://secure-nikeplus.nike.com/plus/
[6] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Berlin, Germany: Springer, 2009, pp 1–4.
[7] P. Bharti, A. Panwar, G. Gopalakrishna, and S. Chellappan, "Watch-dog: Detecting self-harming activities from wrist worn accelerometers," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 3, pp. 686–696, May 2018.
[8] G. Blumrosen, B. Fishman, and Y. Yovel, "Noncontact wideband sonar for human activity detection and classification," *IEEE Sens. J.*, vol. 14, no. 11, pp. 4043–4054, Nov. 2014.
[9] G. Blumrosen, Y. Miron, N. Intrator, and M. Plotnik, "A real-time kinect signature-based patient home monitoring system," *Sens.*, vol. 16, no. 11, 2016, Art. no. 1965.
[10] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surveys*, vol. 46, no. 3, 2014, Art. no. 33.
[11] Q. Chen, B. Tan, K. Chetty, and K. Woodbridge, "Activity recognition based on micro-doppler signature with in-home Wi-Fi," in *Proc. IEEE 18th Int. Conf. E-Health Netw. Appl. Serv.*, 2016, pp 1–6.
[12] J. Cheng, O. Amft, and P. Lukowicz, "Active capacitive sensing: Exploring a new wearable sensing modality for activity recognition," in *Proc. Int. Conf. Pervasive Comput.*, 2010, pp. 319–336.
[13] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. Lamarca, L. Legrand, A. Rahimi, A. Rea, G. Bordello, B. Hemingway, P. Klasnja, K. Koscher, J. Landay, J. Lester, D. Wyatt, and D. Haehnel, "The mobile sensing platform: An embedded activity recognition system," *IEEE Pervasive Comput.*, vol. 7, no. 2, pp. 32–41, Apr. 2008.
[14] *CRFSharp*, 2013. [Online]. Available: http://crfsharp.codeplex.com/
[15] D. Curone, E. L. Secco, A. Tognetti, G. Loriga, G. Dudnik, M. Risatti, R. Whyte, A. Bonfiglio, and G. Magenes, "Smart garments for emergency operators: The proetex project," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 694–701, May 2010.
[16] D. De, P. Bharti, S. K. Das, and S. Chellappan, "Multi-modal wearable sensing for fine-grained activity recognition in healthcare," *IEEE Internet Comput.*, vol. 19, no. 5, pp. 26–35, Sep./Oct. 2015.
[17] S. Gaglio, G. L. Re, and M. Morana, "Human activity recognition process using 3-D posture data," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 5, pp. 586–597, Oct. 2015.
[18] *Gimbal Bluetooth Beacon*, [Online]. Available: http://www.gimbal.com/gimbal-beacons/
[19] P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single triaxial accelerometer," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1780–1786, Jun. 2014.
[20] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp 359–366.
[21] M. A. Hall, "Correlation-based feature selection of discrete and numeric class machine learning," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 359–366.
[22] M. Hassan, W. Hu, G. Lan, S. Khalifa, A. Seneviratne, and S. K. Das, "Kinetic-powered wearable iot for healthcare: Challenges and opportunities," *IEEE Comput.*, to appear, 2018.
[23] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.
[24] A. Jalal, S. Kamal, and D. Kim, "Shape and motion features approach for activity tracking and recognition from kinect video camera," in *Proc. IEEE 29th Int. Conf. Adv. Inf. Netw. Appl. Workshops*, 2015, pp 445–450.
[25] T.-P. Kao, C.-W. Lin, and J.-S. Wang, "Development of a portable activity detector for daily activity recognition," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jul. 2009, pp 115–120.
[26] S. Khalifa, M. Hassan, A. Seneviratne, and S. K. Das, "Energy-harvesting wearables for activity-aware services," *IEEE Internet Comput.*, vol. 19, no. 5, pp. 8–16, Sep. 2015.

[27] S. Khalifa, G. Lan, M. Hassan, A. Seneviratne, and S. K. Das, "HARKE: Human activity recognition from kinetic energy harvesting data in wearable devices," *IEEE Trans. Mobile Comput.*, vol. 17, no. 6, pp. 1353–1368, Jun. 2018.

[28] A. Khan, Y.-K. Lee, S. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1166–1172, Sep. 2010.

[29] I. Kononenko, "Estimating attributes: analysis and extensions of relief," in *Proc. Eur. Conf. Mach. Learn.*, 1994, pp. 171–182.

[30] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," in *Proc. Conf. Emerging Artif. Intell. Appl. Comput. Eng.: Real Word AI Syst. Appl. eHealth HCI Inf. Retrieval Pervasive Technol*, 2007, pp. 3–24.

[31] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn.*, 2001, pp. 282–289.

[32] O. Lara and M. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, Jul.-Sep. 2013.

[33] J. W. Lockhart, T. Pulickal, and G. M. Weiss, "Applications of mobile activity recognition," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 1054–1058.

[34] J. W. Lockhart and G. M. Weiss, "The benefits of personalized smartphone-based activity recognition models," in *Proc. SIAM Int. Conf. Data Mining*, 2014, pp. 614–622.

[35] D. L. Mills, "Internet time synchronization: The network time protocol," *IEEE Trans. Commun.*, vol. 39, no. 10, pp. 1482–1493, Oct. 1991.

[36] K. P. Murphy, *Naive Bayes Classifiers*, Vancouver, BC, Canada: Univ. British Columbia, 2006.

[37] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors, " *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, Jan. 2006.

[38] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, pp. 1883, 2009.

[39] J. R. Quinlan, "Improved use of continuous attributes in c4. 5," *J. Artif. Intell. Res.*, vol. 4, pp. 77–90, 1996.

[40] D. Riboni and C. Bettini, "COSAR: Hybrid reasoning for context-aware activity recognition," *Personal Ubiquitous Comput.*, vol. 15, no. 3, pp. 271–289, 2011.

[41] N. Roy, C. Julien, A. Misra, and S. K. Das, "Quality and context-aware smart health care: Evaluating the cost-quality dynamics," *IEEE Syst. Man Cybern. Mag.*, vol. 2, no. 2, pp. 15–25, Apr. 2016.

[42] N. Roy, A. Misra, and D. J. Cook, "Infrastructure-assisted smartphone-based ADL recognition in multi-inhabitant smart environments," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2013, pp. 38–46.

[43] S. Saguna, A. Zaslavsky, and D. Chakraborty, "Complex activity recognition using context-driven activity theory and activity signatures," *ACM Trans. Comput.-Human Interaction*, vol. 20, no. 6, pp. 32:1–32:34, Dec. 2013.

[44] *Samsung Galaxy S4*, [Online]. Available: http://www.samsung.com/us/mobile/cell-phones/SGH-M919ZWATMB

[45] J. T. Townsend, "Theoretical analysis of an alphabetic confusion matrix," *Attention Perception Psychophysics*, vol. 9, no. 1, pp. 40–50, 1971.

[46] P. Vepakomma, D. De, S. K. Das, and S. Bhansali, "A-wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities," in *Proc. IEEE 12th Int. Conf. Wearable Implantable Body Sens. Netw.*, Jun. 2015, pp. 1–6.

[47] L. Wang, T. Gu, X. Tao, H. Chen, and J. Lu, "Recognizing multi-user activities using wearable sensors in a smart home," *Pervasive Mobile Comput.*, vol. 7, no. 3, pp. 287–298, Jun. 2011.

[48] G. Wolf, "The quantified self". [Online]. Available: http://antephase.com/quantifiedself

[49] J. B. Murphy, W. D. Spector, S. Katz, and J. P. Fulton, "The hierarchical relationship between activities of daily living and instrumental activities of daily living," *J. Chronic Diseases*, vol. 40, no. 6, pp. 481–489, 1987.

[50] D. H. Wilson and C. Atkeson, "Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors," in *Proc. 3rd Int. Conf. Pervasive Comput.*, 2005, pp. 62–79.

[51] X. Yang and Y. Tian, "Super normal vector for human activity recognition with depth cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 1028–1039, May 2017.

[52] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

[53] K. Zhan, S. Faux, and F. Ramos, "Multi-scale conditional random fields for first-person activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2014, pp. 51–59.

[54] Z. Zhao, Y. Chen, J. Liu, Z. Shen, and M. Liu, "Cross-people mobile-phone based activity recognition," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 2545–2550.

[55] C. Zhu and W. Sheng, "Human daily activity recognition in robot-assisted living using multi-sensor fusion," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2009, pp. 2154–2159.

**Pratool Bharti** received the PhD degree in computer science and engineering from the University of South Florida. He is a research and development manager. His research interest include smart healthcare, computer vision, and deep learning.

**Debraj De** received the PhD degree in computer science from Georgia State University. He is CTO and a principal engineer with Smart Health Beacons LLC. His research interest includes smart healthcare.

**Sriram Chellappan** received the PhD degree in computer science and engineering from Ohio State University. He is an associate professor with the Department of Computer Science and Engineering, University of South Florida. Previously, he was a faculty member with the Department of Computer Science, Missouri University of Science and Technology. His current research interests include social and technological aspects in smart healthcare, privacy, and cybersecurity.

**Sajal K. Das** received the PhD degree in computer science from the University of Central Florida. He is a professor of computer science and Daniel St. Clair Endowed chair with the Missouri University of Science and Technology and also a distinguished visiting professor with Zhejiang Gongshang University, Hangzhou, China. His current research interests include wireless sensor networks, smart healthcare, cyber-physical systems, mobile and pervasive computing, security and privacy, and social networks. He is a fellow of the IEEE.