AI at the Disco

Low Sample Frequency Human Activity Recognition for Night Club Experiences

Amritpal Singh Gill*
Huawei
Amsterdam, The Netherlands
amritpal.singh.gill@huawei.com

Sergio Cabrero*
IBM
Dublin, Ireland
sergio.cabrero.barros@ibm.com

Pablo Cesar†
Centrum Wiskunde & Informatica
Distributed and Interactive Systems
Amsterdam, The Netherlands
P.S.Cesar@cwi.nl

David A. Shamma
Centrum Wiskunde & Informatica
Distributed and Interactive Systems
Amsterdam, The Netherlands
aymans@acm.org

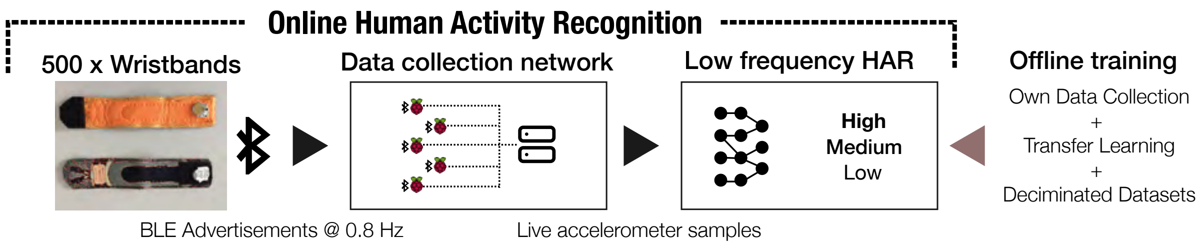


Figure 1: System diagram of a low frequency Human Activity Recognition

ABSTRACT

Human activity recognition (HAR) has grown in popularity as sensors have become more ubiquitous. Beyond standard health applications, there exists a need for embedded low cost, low power, accurate activity sensing for entertainment experiences. We present a system and method of using a deep neural net for HAR using low-cost accelerometer-only sensor running at 0.8Hz to preserve battery power. Despite these limitations, we demonstrate an accuracy at 94.79% over 6 activity classes with an order of magnitude less data. This sensing system conserves power further by using a connectionless reading—embedding accelerometer data in the Bluetooth Low Energy broadcast packet—which can deliver over a year of human-activity recognition data on a single coin cell battery. Finally, we discuss the integration of our HAR system in a smart-fashion wearable for a live two night deployment in an instrumented night club.

*Work performed author was at Centrum Wiskunde & Informatica.
†Also with Delft University of Technology.

HuMA'20, October 12, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-8151-2/20/10…$15.00
https://doi.org/10.1145/3422852.3423485

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Computer systems organization** → **Sensors and actuators**; • **Applied computing** → *Health informatics*; • **Computing methodologies** → Neural networks.

KEYWORDS

HAR, human activity recognition, deep learning, CNN, RF, low-frequency, sampling, power, battery, nightclub

ACM Reference Format:
Amritpal Singh Gill, Sergio Cabrero, Pablo Cesar, and David A. Shamma. 2020. AI at the Disco: Low Sample Frequency Human Activity Recognition for Night Club Experiences. In *1st International Workshop on Human-centric Multimedia Analysis (HuMA'20), October 12, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3422852.3423485

1 INTRODUCTION

The past decade have seen a rapid growth in the field of microelectronics and computer systems; sensors and mobile devices are getting smaller, cheaper, and more powerful, and thus have become an important part of our daily lives [18]. This has given rise to new methods in ubiquitous computing to extract and process the sensor data in embedded devices [23]. *Human Activity Recognition* (HAR) is a dominant issue that aims at determining activities, like running, walking, and standing, based on the data acquired from relevant sensors [30, 36]. HAR finds its applications in a wide variety of domains from healthcare to security to sports and entertainment. For instance, HAR covers a range of utilities from health monitoring [34], to self-awareness [9], to game interaction [11]. Many of these examples use rechargeable higher cost devices with high frequency sampling. However, there are some smart place and entertainment driven applications where a lower cost device with low frequency sampling is ideal.

In this paper, we present a method for HAR when the frequency of accelerometer samples is low. Our approach uses a *Convolutional Neural Network* (CNN); we have developed and validated it using publicly available HAR heath tracking datasets. We then apply it in a scenario with a high number of concurrent *Bluetooth Low Energy* (BLE) enabled devices that run on a small coin battery for weeks. Unlike most existing systems, our goal is not directly related to personal informatics; we build an interactive physical space, a nightclub, with hundreds of participants whose activity level while dancing influenced aspects of the space. A low power lightweight sensor built into a fashion bracelet communicates with an environment which we also instrumented for a large dance event.

In conjunction with two nights of the *Amsterdam Dance Event* festival of 2016, averaging 450 guests a night, the club was a curated and designed experience to stimulate all the senses at once: specially created dinner menus, top DJs, drinks and perfumes, an adaptive sound system, and projected visualizations. Our sensing technology had to play the role of connecting people to the club by monitoring their activity level (dancing, walking, or standing) and where they were in the club.

As we were targeting hundreds of users a night in a small set of confined spaces, we used low broadcast frequency devices to prolong battery life and to reduce the amount of interference among wristbands; higher broadcasting frequencies lead to higher battery consumption and increase congestion in the 2.4GHz band which can cause high packet loss ratio with our expected density of devices. However, a low sampling frequency solves battery and noise issues but makes HAR difficult. Typical sensors, accelerometers and gyroscopes, in HAR systems collect data at high enough sampling frequency to capture relevant body movements. Most research claims that a sampling frequency of at least 20Hz can properly assess human activities [20]. Attempts to overcome this battery life limitation have employed algorithms that dynamically adapt different sampling frequencies based on predefined heuristics [7, 19, 20, 32]. Though these attempts have shown low sampling frequencies (< 5 Hz) to reduce battery consumption by over 50%, their operation is still dependent upon the partial use of high sampling frequencies. Based on non-neural network machine learning models, these studies employ handcrafted extracted features based on an expert or domain knowledge with high frequency sensor sampling. Modern CNNs allow for automatic feature learning from data [1, 13, 22, 27, 33, 35], and have outperformed previous handcrafted feature based methods.

Like other CNN approaches, there is no need to handcraft features. In addition, we obtain comparable performance at 94.79% accuracy (2% less than the state of the art approach) using only triaxial accelerometer readings operating at a 1Hz sampling frequency—an order of magnitude less than other approaches. We then adapt the activity classifier for a night club environment through transfer learning to predict three classes: standing, walking, and dancing. Finally we describe the system's integration into the overall club experience.

2 RELATED WORK

Activity recognition has been growing along with the rise of personal tracking and low cost well-being sensing. One must consider what sample data, portability, data extraction, online vs. offline processing, real environments, and power consumption. When these considerations couple with interactions and experiences around HAR the sensor device, overall system, and application becomes unique. Even how one wears a sensor has an effect in what we can measure. [29] These applications and systems are often driven from a combination of Artificial Intelligence, Internet of Things, and real-time analysis.

2.1 Personal Wearable Activity Sensors

In recent years, wearable technology is empowering people to quantify themselves (e.g., FitBit[1], Withings[2], and Empatica[3]). These devices allow logging and tracking daily activities, by capturing relevant sensor data which is later aggregated and visualized to the user and in some cases to other users and relatives. Relevant application domains include sport training, well-being, and health support. In terms of technology, these devices are normally paired with the mobile phone, include some storage capabilities, and are rechargeable. This allows for high-frequency sampling as consumer devices require weekly charging if not daily. For experiments, battery life poses an issue; one experiment even duct-taped a battery pack to a device [4].

Bentley et al. published an extensive article about well-being applications [3]. They developed a working system that combined inputs (self-reports, mobile tracking and wearable sensing) and provided to the users richer and significant observations in a mobile application. By providing significant information, displayed in a natural manner, users were more engaged. Moreover, there was an increase on self-awareness and reflection, which lead to changes in behavior in their daily lives. Overall, the article provides insights on better visualization mechanisms of self-data based on aggregation of sensor data.

An analysis of Quantified Self Meetup video talks [8] provide insights about the motivations behind self-tracking, the tools that are commonly used and the self-reflection of the users based on the tracked data. According to the results, future systems should provide mechanisms to track the context in which the captured data to ease its interpretation. Value-Sensitive Design [12] can too address privacy, visualization, and comprehension in HAR systems and applications. It demonstrates that a binary choice for sharing is not adequate and the need for richer and more complex mechanisms. In particular, the authors recommend to provide support for intermediate levels of sharing and mechanisms to control over-sharing. Picard [24] offers a reflection about the process of taking sensors from the lab to the wild. Their emphasis is on wearable devices for health applications, which can alert of potentially dangerous situations like epileptic seizures.

Along with research on daily activity tracking, the contribution of the work is that our HAR system robustly operates under

[1]https://www.fitbit.com/nl/home
[2]https://health.nokia.com/nl/en/
[3]https://www.empatica.com

low-frequency sampling. Our main motivation is lower power consumption of the devices so they can last for the longest period (ideally months) of time without recharging. This allows the devices that do not need recharging, which can increase the price and the size, and in some cases (as with smart clothes) may affect the aesthetics. The second motivation is the system and application domain. Existing devices are typically for self-monitoring, only requiring paring with the mobile phone of the owner, which is often close by. We are looking instead into communication scenarios, where the sensors communicate in real-time with the environment and ultimately with others under noisy conditions. Such conditions naturally decrease the de-facto available sampling frequency and introduce a random effect in it, as eventually the samples received will depend on factors like the distance to the receiver or number of interfering transmission devices.

2.2 Public Datasets

Sensing and sensing data typically uses a combination of accelerometer, gyroscope, and magnetometer and conducted at a rate of 30Hz or greater. Additionally, these datasets also may include multiple sensors mounted at multiple places on the body. Our goal is to use a single accelerometer only sensor at approximately one sample per second (1Hz) to maximize battery life on a small, lightweight embedded sensor. Table 1 outlines the HAR datasets we discuss here along with prediction methods and accuracy scores.

2.2.1 WISDM. The WISDM (Wireless Sensor Data Mining) [21] dataset uses smart phone-based sensor readings. It details three main design issues: (1) devices have limited resources (power, computational speed, and bandwidth), (2) mining must scale to thousands or millions of users, and (3) results must be, at minimum, categorized in real-time. The resulting dataset records 6 daily activities performed by 563 subjects in a real world scenario, recorded using a tri-axial accelerometer equipped smart-phone placed in their pockets. The recorded activities includes walking, jogging, climbing stairs, sitting, standing and lying down. The dataset uses a 20Hz sampling speed and contains in total 2,980,765 data points.

2.2.2 SmartLab Version 1. The SmartLab dataset for HAR (version 1) [2] was created to unobtrusively monitor activities for daily living. This dataset records 6 daily activities performed by 30 subjects in a controlled environment, recoded using a tri-axial accelerometer and a gyroscope embedded in a Samsung Galaxy SII smart-phone fixed on the waist of the participants. The recorded activities are walking, walking upstairs, walking downstairs, sitting, standing and laying. This dataset, collected at 50Hz sampling speed, contains in total 747,550 data points. When pre-processing the data, the authors used a median filter and a 3rd order low-pass Butterworth filter cutoff at 20Hz, as they cite that 99% of body motion energy is below 15Hz [15]. A second version of this dataset [26] included postural transitions between activities which was not utilized in our studies here as transitions were not our focus.

2.2.3 USC-HAD. University of Southern California Human Activity Dataset (USC-HAD) [36] records 12 daily activities performed by 14 subjects in a controlled environment, recoded using a tri-axial accelerometer and a gyroscope fixed on the waist of the participants. The recorded activities are walking forward, walking left, walking right, walking upstairs, walking downstairs, running forward, jumping, sitting, standing, sleeping, elevator up, elevator down. This dataset, collected at 100Hz sampling frequency, contains in total 2,811,490 data points. USC-HAD focused on having a diverse set of participants (age, gender, weight, and height) with a specialized fixed mounted sensor to capture accurately and robustly; the authors used an off the shelf MotionNode (wired sensor) as it captured at a higher sampling rate than smartphones at the time but was still unobtrusive.

2.3 HAR Methods and Low Hz Adaptations

Not untypical for *Machine Learning* (ML), there were a variety of techniques over the years for HAR classification including *k-nearest neighbors* (k-NN), *Random Forest* (RF), *Support-Vector Machines* (SVMs), and later CNNs. With RF [6] achieving above 90% accuracy using a private dataset. In 2014, high accuracy using CNNs [14] was also shown for HAR. Beyond the ML methods, we focus on the mechanics of adjusting the sampling frequency of the sensor to conserve power.

Much of the research, including datasets in HAR, describe battery-life and power as a design constraint; indeed there is a trade-off between energy and accuracy. In some cases, like USC-HAD [36], it ignores power constraints with the general assumption that battery issues are a different issue. However, other research does address how to improve power consumption but yet retain accuracy. While many of the low-frequency adaptations use on-device HAR computation, our aim is to have a sensor report the data elsewhere and compute the recognition elsewhere.

2.3.1 Adaptive frequency. As it is the focus of our research, adjusting the sampling frequency is the first step in power-conservation. Since most activities require around 15Hz [15], high resolution sampling is generally extraneous. When coupled with a hierarchical recognition scheme, a recognition accuracy of 85% is possible and battery life extended for hours on a smartphone [20].

2.3.2 Adaptive frequency and window size. Along with sampling frequency, the *window size* (WS) for samples is also adjustable. Different activities carry different typical duration (e.g. we tend to walk for a longer period of time than we would climb stairs), but generally lowering the sampling frequency while increasing the WS can maintain accuracy in HAR. When done dynamically, it is optimal to inversely adjust sampling frequency and WS once an activity is recognized [19] (the assumption being that activity state will persist for a short duration).

2.3.3 Adaptive features and other techniques. Additionally, there are other optimizations strategies. Typically, the classification features are also dynamically set to save battery life [19, 32]. Likewise, probability models looking for the likelihood of an activity's occurrence coupled with a multi-class *Support-Vector Machine* (SVM) can save approximately half the energy at 100Hz [7]. These techniques show significant power savings for on device recognition. Davila et al. [10] propose an iterative learning framework with different filtering mechanisms to reduce the number of samples needed for HAR. Their method reduces training times by selecting only the samples that are most relevant.

Table 1: Three baseline datasets used for *Human Activity Recognition* with prediction scores for *Random Forest* (RF) and *Convolutional Neural Networks* (CNNs) methods: SmartLab-v1 [2], USC-HAD [36], WISDM [21]. Input: A = Accelerometer, G = Gyroscope. Features: h = hand crafted, t = time, f = frequency.

Algorithm	Sensor	Rate	Dataset	Features	Accuracy
RF [6]	A	50Hz	Private	h	94.00%
CNN [27]	A, G	50Hz	SmartLab-v1	t	94.79%
CNN [27]	A, G	50Hz	SmartLab-v1	t, f	95.75%
CNN [13]	A, G	50Hz	SmartLab-v1	f	97.59%
CNN [13]	A, G	100Hz	USC-HAD	f	97.83%
CNN [35]	A	20Hz	WISDM	t	96.88%

Table 2: A table expected battery life of an Estimote by Advertising Interval and Broadcasting Power.

	Broadcast Power		
Advertising Interval	-30 dBm	-4 dBm	+4 dBm
2000 ms	3.3 years	3 years	2.3 years
1000 ms	1.9 years	1.7 years	1.3 years
600 ms	1.2 years	1 year	300 days
200 ms	160 days	140 days	104 days
50 ms	40 days	35 days	26 days

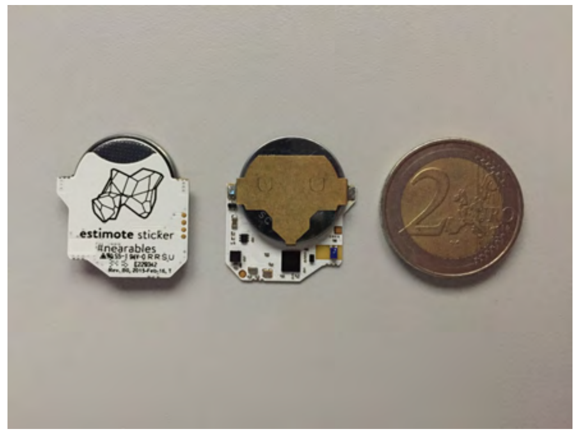
3 METHOD

We present two models, RF and CNN, for *Human Activity Recognition* that operate under our low power, multi-day constraints. First, we discuss our target device for development. Next, we create two models using health tracking data. While our final application is entertainment driven, we will use health tracking benchmarks as our testing standard as they are more rigorous. First, we replicate a known RF Model then we implement a novel CNN to better handle low sampling frequency.

3.1 Devices and Data

While the HAR datasets [2, 17, 36] use high sampling frequency sensors,we will be using an off-the-shelf Estimote sensor. The Estimote carries many limitations and retails for about 10 USD. It is a small, easily embeddable, coin-battery operated BLE chip with a accelerometer and temperature sensor on board. At a broadcasting power of +4 dBm, the battery lasts approximately 1.3 years, which is ample time to run a longer study or embed into a device or clothing without demanding constant recharging. The Estimote bundles the sensor data into the BLE broadcast packet—an advertisement—which allows connection-less reading. BLE's Advertisement Interval range is, by specification, 20ms–10.24s plus a random delay of 0–10ms. A packet is sent three times across various channels. So the minimum delay between packets of *new information* is 60–90ms. Consequently, the broadcast rate of 1Hz is effectively about 0.8Hz, even lower if interference in the broadcast channel exists, and is without evenly spaced delivery. While it is possible to adjust the advertising rate, at 20Hz and +4dBm[4] the anticipated battery life, from Table 2, is about 26 days. Second, the available datasets used a plurality of sensors including gyroscope and magnetometer. For our purposes, we will only use the accelerometer data from these datasets in our experiments (which is a limitation of the Estimote sensor). We will also test predictions at 0.5Hz and 1Hz. As the datasets vary from 20–100Hz sample rates, we undersample them

[4]http://blog.estimote.com/post/83618039493/how-to-extend-estimote-beacon-battery-life



(a) An Estimote *Bluetooth Low Energy* sensor (printed circuit board only) with 2€ coin for scale.

(b) The bespoke wristband made using a circular knitting machine. Overlaid Estimotes show the placement of the embedded sensor.

Figure 2: The Estimote sensor and circular knit wristband.

accordingly by removing every n samples at a regular interval to achieve the slower 0.5Hz and 1Hz rates.

3.2 Random Forest

In our first approach, we do prediction via Random Forest. Here the accelerometer data from the SmartLab-v1 dataset, using a 50% overlapping sliding windows. The size of the sliding window corresponded to the used sampling rate where 50Hz would yield 500 samples over 10 seconds and 1Hz would yield 10 samples over 10 seconds. Per instance of each sliding window, we extracted the following 11 features: (i) Mean, (ii) Standard Deviation, (iii) Range ($t_{\text{Max}} - t_{\text{Min}}$), (iv) Root mean square, (v) Zero crossing rate of amplitude, (vi) Amplitude kurtosis, (vii) Argmax of Fourier transform, (viii) Max of Fourier transform, (ix) Mean of Fourier transform, (x) Standard deviation of Fourier transform, and (xi) Kurtosis of Fourier transform, With these features, 10 fold cross validation was used to train the random forest classifier using the computed features, with 500 estimators. Cross validation can be done by dividing data based on the subjects from which the data were collected. This is *subject-sensitive* cross validation. Another way of performing cross validation is by dividing the data received from every subject into folds such that each fold contains data received from all of the subjects. This is *activity-based* cross validation. The latter generally performs at a higher accuracy yielding 93.46% @ 50Hz, 70.67% @ 1Hz, and 71.06% @ 0.5Hz.[5] While high frequency sampling performed well, we lost performance at under 1Hz rates using RF.

3.3 Convolutional Neural Network

Having completed a baseline Random Forest experiment, we focus on building a CNN for real-time HAR classification. Recent advancements in deep learning have demonstrated the efficacy of this approach in HAR [31]; further the use of a CNN with a SoftMax filter provides multi-class classification of a single model. While these approaches use the accelerometer and gyroscope data together at 50–100Hz [13, 27], there have been some research at a lower frequency (20Hz) using the accelerometer only. [35] Table 1 shows a accuracy comparison of various techniques where the 20Hz accelerometer only prediction [35] reached an accuracy of 96.88% (μ = 96.60% across all techniques).

For our CNN, we will use both the SmartLab-v1 dataset and the WISDM dataset in two individual trials. As before, will only observe the accelerometer data and we will set the sliding window, corresponding to the sample rate used the same as we did in the random forest model. Here, we use an overlapping window sliding by one data point, over the time series accelerometer data. We use two convolution layers with ReLU activation function, and both with filter size of 5 and 32 filters per layer. Next is a max pooling layer with a pool size of 2 and then a *Long Short Term Memory* (LSTM) layer with 32 units in total. Then we add a dropout layer with a probability of 0.8. Finally, we add a dense layer with SoftMax activation function for the classification. Table 3 shows the overview of the CNN layers. We used Adam [16] optimizer for training our CNN, and used minibatch gradient descent with minibatch size of 16 and a 0.02 learning rate.

[5]See supplemental Table A.1 for expanded results.

Table 3: The architecture of our CNN.

Layer	Type	Output Shape	Params
1	Conv1D	(None, 10, 32)	512
2	Conv1D	(None, 10, 32)	5,152
3	MaxPooling1D	(None, 5, 32)	0
4	LSTM	(None, 32)	8,320
5	Dropout	(None, 32)	0
6	Dense	(None, 6)	198
			Trainable params: 14,182
			Non-trainable params: 0

Our results show success with the undersampled SmartLab-v1 dataset utilizing only the accelerometer and our sliding window and our CNN. We achieve an accuracy of 94.7% at 1Hz and at 0.5 Hz.[6] This is particularly needed as BLE advertising packets are susceptible to loss (discussed further in Section 4); so efficiency at 0.5Hz exhibits a tolerance to packet loss. This is approximately 3% less than state of the art which used an additional gyroscope and higher sampling. The WISDM dataset in the same CNN fared worse losing 7.38% accuracy at 1Hz.[7] The *Random Forest* method lost the most performance at a lower sampling frequency (almost 22%). The difference between the RF and the CNN demonstrates the need for a deep network for this task. While each dataset has some variability based on placement and sensor/sensors, overall we find that we can predict above 94% with an order of magnitude less data. However, *it comes with a limitation* of response time; our method requires 10–20 seconds of data collected to make a prediction. In effect, this can take close to 30 seconds to detect an activity change (like running to walking).

4 FIELD STUDY DEPLOYMENT

We have shown how a low cost, low power, low sampling frequency sensor can deliver a year of *Human Activity Recognition* data for prediction at 94.7% accuracy. This saves the need to recharge or replace batteries and allows for light-weight long-life smart-wearables without added bulk or weight. Non-health tracking and HAR applications, one of our motivations, is fashion wearables for smart nightclub entertainment environments. This requires some new systems infrastructure along with transfer learning our activity recognition CNN into a new domain.

4.1 Wristband Considerations

The combination of technical requirements, available time, and resources created a complex set of constraints. Both the individual and the collective experience were important. This includes the guests' privacy. They should not feel tracked and the system should not put a burden on them; they were in the party to have fun. Others have used video as a possible solution to solve monitoring individuals in a crowd [18, 28], but it also raises new privacy concerns in this context. Thus, body-worn sensors were the best (least surveillance driven) alternative. Contrary to the popular approach of pairing wearables and smartphone apps, the solution should not

[6]See supplemental Table A.2 for expanded results.
[7]See supplemental Table A.3 for expanded results.

Table 4: A comparison of methods by datasets and sampling frequency. Our method is comparable to similar time-based methods but approximately 3% less than time and frequency based methods. Note how the RF was not robust under degradation; this demonstrates why a deep network is required for this task. Our best results in bold for each algorithm. Datasets: SmartLab-v1 [2], USC-HAD [36], WISDM [21]. Sensor: A = Accelerometer, G = Gyroscope. Features: h = hand crafted, t = time, f = frequency.

Algorithm	Method	Sensor	Rate	Dataset	Features	Accuracy
RF	Casale et al. [6]	A	50Hz	Private	h	94.00%
	Our Method	A	50Hz	SmartLab-v1	h	93.46%
		A	1Hz	SmartLab-v1	h	**71.67%**
		A	½ Hz	SmartLab-v1	h	71.06%
CNN	Ronao and Cho [27]	A, G	50Hz	SmartLab-v1	t	94.79%
	Ronao and Cho [27]	A, G	50Hz	SmartLab-v1	t, f	95.75%
	Jiang and Yin [13]	A, G	50Hz	SmartLab-v1	f	97.59%
	Jiang and Yin [13]	A, G	100Hz	USC-HAD	f	97.83%
	Zeng et al. [35]	A	20Hz	WISDM	t	96.88%
	Our Method	A	1Hz	SmartLab-v1	t	**94.79%**
		A	½ Hz	SmartLab-v1	t	94.76%
		A	1Hz	WISDM	t	89.38%
		A	½ Hz	WISDM	t	89.52%

require any effort by the guests or any intrusion in their personal devices. After exploring different options with respect to design and available technology, we fitted bespoke designer wristbands with off-the-shelf, BLE-enabled circuit boards. Using BLE advertisement broadcast messages and a network of Raspberry Pis strategically located in every room as sniffers, we were able to minimize deployment and setup times. The night club would pickup guests' accelerometer data from their wrist bands when in reach. Just by wearing their wristbands, guests could take part of the experience. These design choices lead us to low frequency sampling for each individual wrist bands due to: battery life and data rate available with a high number of devices sharing the BLE broadcast channel.

Several prototypes tested both for aesthetic look and sensor placement. We opted for the design shown in Figure 2b made in conjunction with a fashion designer using a circular knitting machine. This style of machine allows for quilted pockets in the construction that are ideal for small embeddable sensors. Visually, the design was closer to a *sophisticated* soft sweatband opposed to a modern plastic fitness tracker. Thus, the embedded sensor was invisible and unnoticeable. While the guest all had informed consent, the goal was to make them *mostly* forget about the wearable.

We needed to create a system that could collect data in real-time in order to run the HAR classification, but it also had to fulfill our requirements. It would be too much overhead to have each patron receive a wristband, download an app, and pair it to a smart phone. As an alternative, we informed the night club of the sensors that would be broadcasting in the environment. This meant 900 sensors needed server registration, then embedded in the partially constructed wristbands, and finally sealed inside the completed wristbands. Even at 20Hz, the Estimotes would last the two weeks for this process to complete. In this case our power saving method gave us months of time to use the devices (and could easily be reused months later by a patron for a longer multi-week experiment).

Figure 3: Setup of the night club and the RaspberryPI actuators and receivers. The numbers listed are the ID of the device.

4.2 Room system architecture

Given these requirements, we chose to collect data through *Bluetooth Low Energy* (BLE) advertisement packets—the protocol used by Apple's *iBeacons*. While we produced two wristband models with different embedded sensors, we focus on the 800 Estimote-powered wristbands. The other 100 bands used a *SensorTag CC2650* board from *Texas Instruments (TI)*. It is larger than the Estimote board, but it is a more general-purpose board for IoT applications, has more sensors built-in and is fully programmable. Even throttled to 0.8Hz BLE broadcast, these lasted only two days on a single coin cell.

No sensor is ever connected to and all the accelerometer data are within the BLE advertisements which were collected via a network of Raspberry Pis throughout the venue. This connection-less model for data collection has advantages and disadvantages. On the positive side, it allows one to monitor packets without a formal connection (saving power). Additionally, the implementation is straightforward. Raspberry Pis just listen to BLE advertisements of wristbands in range, and forward them to the server. There is no need to implement any roaming mechanism, which would be necessary in a connection oriented scheme. Moreover, the overhead stays minimum—no connections have to be maintained—so sensors have a simple logic and dedicate their energy only to send data. On the other hand, the main disadvantage is that the channel is unreliable: if there is data loss, it is lost forever. Although a BLE advertisement containing the same data sample is sent three times through three different channels, chances are that the wristband is out-of-range of all receivers or that all three packets are lost due to noise. The latter is more likely in environments with high density of people and devices. To cope with this, we deployed a high number of Raspberry Pis in the space, which increases the chances of receiving each data sample at least once (see Figure 3). A collateral effect is that the same data sample is sometimes received multiple times by different Raspberry Pi, so we implemented a filter for duplicated samples. We also moved the WiFi network inside the club to 5GHz only to reduce the utilization of the 2.4GHz band as much as possible. Despite our efforts, losing data samples in such a dense environment is unavoidable. Thus, our HAR system had to cater for lost data samples that in practice implies a variable and not deterministic sampling frequency. For some wristbands and in some moments of the event, we observed a sampling frequency lower than half the expected 0.8Hz.

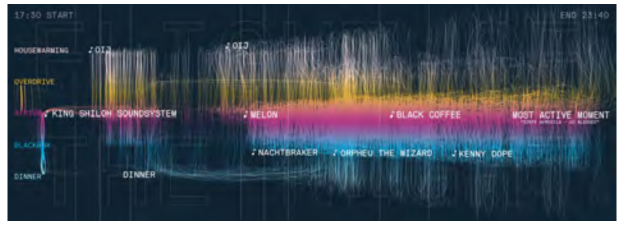
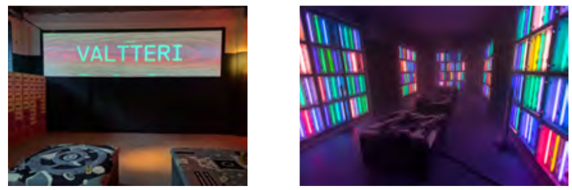


Figure 4: A patron's visualization of their night. Graph shows activity level over time on the x-axis and room (housewarming, overdrive, atrium, blackbox, dinner) on the y-axis. Moments of prolonged dancing are highlighted with the DJ and/or song that was playing at that time.



(a) **(b)**

Figure 5: Photos of responsive elements in the nightclub. The reception (a) would rotate usernames and dance activity levels in a large public display. The lounge's (b) lights and sound reflected the inverse HAR energy in the club.

4.3 Transfer Learning

For the nightclub, we had to predict three motion classes: standing, walking, and dancing. The data for these classes were collected privately in our institution. For standing class, the data were collected by placing a number of sensors in a box and recording their accelerometer readings while giving small jitters to the containing box. Since in an uncontrolled environment like that of a nightclub, the guests usually do not stay perfectly idle while standing, small jitters during data collection helped in collecting robust data points that are closer to the real scenario. For walking class, a number of volunteers from our institution wore the sensor embedded in a wrist sweat/tennis band. They were all instructed to walk through the halls at different paces. For the dance data, we hosted two 30-minute dance "parties" with our institution participants. Some participants danced energetically while others simply swayed in place holding a beverage. There were 6 participants in total (4 male, 2 female) aged between 24 and 44 years old.

The limited amount of our collected data was not sufficient to train our CNN (§ 3.3) without risking model over-fitting. To ensure robustness, better generalization, and to avoid over-fitting, we used *transfer learning* in a similar manner as done in computer vision [25] tasks. In transfer learning, high-order and complex correlations existing in the input data are first learned by pretraining a CNN using a sufficiently big dataset. Features learned from pretraining are then fine-tuned to the requirements of the end-task. The end-task does not require a lot of data for fine-tuning because most of the relevant features are already learned in the pretraining step. In this manner learning is transferred from pretraining to fine-tuning.

Following transfer learning, we used the SmartLab-v1 [2] dataset for pretraining our CNN to predict six motion classes. Since our nightclub application required only three motion classes, we re-initialized the dense layer (#6 in Table 3) of our CNN to predict three classes instead. All other layers and weights were retained from our pretrained CNN. At last, we fine-tuned our CNN using our collected dataset to achieve approximately 95% classification accuracy on the three classes required for the nightclub experiment.

4.4 Sensor Localization

Further, we used a BLE localization method (out of scope for this article) to determine which one of the five rooms (housewarming, overdrive, atrium, blackbox, dinner) each patron was in over the course of each night. Together, the localization with the activity recognition enabled us to automatically control lighting and effects in the club. Figure 5 shows how the environment was instrumented during the event. Further, days after the event, each patron received a visualization of their night depicting in which rooms they were in over the night and at which times they were dancing for the longest duration along with the song and/or DJ who was playing during that time. See Figure 4 for a visualization of one patron's night. The data collected from the two nights is available for download [5].

5 DISCUSSION

We have identified and demonstrated low frequency HAR for entertainment venues and multimedia installations. We demonstrated a CNN that performs 2% under than state of the art health tracking at an order of magnitude less the sampling rate. This shows we can provide activity detection off-device for a year from a small sensor with single coin cell battery. We then applied this to an entertainment domain set up as a smart place to control interactive media experiences. This used a larger hub server at the center of the edge nodes. The edge nodes (RaspberryPis) themselves can do the evaluation with modern USB GPUs like Intel's NCS2 or Coral's TPU. The overall primary limitation is user activity state changes take longer to recognize at 1Hz vs 60Hz.

5.1 Low power for wearables

We needed about one month of battery life to complete this experiment, we continually checked a sample of the 50 remaining wristbands and found them alive for the full year but only 2 lasted the full predicted 1.3 years. This still shows promise and presents some lessons about using batteries in smart-textiles for HAR and longer run experiments. Future clothes will have embedded sensors distributed in our bodies. Similarly to the application described, these sensors will also require wireless communication and low energy consumption, which is likely to imply low data sampling frequencies. Moreover, a connection-less model enables easier interaction with other computing systems around them, such as other clothes, smart rooms or other people's clothes. For these reasons, we are currently extrapolating this model to reactive clothes that use HAR to modify their physical properties.

5.2 Beyond personal tracking

The existing literature on HAR primarily focuses on the individual and provides mechanisms for tracking activities. This complements research on personal smartphone devices as a gateway between sensors and external cloud services. Our target is different; we focus on measuring individual activity as a way to understand groups and the collective action of those who share an environment. Furthermore, we removed smartphones from the system and used the environment itself as a gateway for the data. As a result, activity monitoring is tightly coupled with the place in which it occurs. Such divergence from the more typical case study results in different assumptions regarding data gathering (noisy environment), type of hardware for sensors, and temporal precision of the recognition.

5.3 Real world scale

We further report the deployment of the solution in a real-world environment during the *Amsterdam Dance Event* festival of 2016. Such real physical deployment (figure 3) is a unique demonstration of our contribution. Apart from complicating the logistics (informed consent, registering almost 1,000 devices, distributing devices to guests), the engineering concerns had the added deployment complications which provide design and architecture considerations for future systems. We demonstrated how to utilizes our HAR system with hundreds of people and devices; moreover, we establish a baseline for this scale of sensing and activity recognition.

5.4 AI and Multimedia Systems

Activity recognition systems, and AI systems in general, are still in early stages of development. Thus, the methods we use cannot assume perfect data gathering which creates additional concerns beyond the typical robust method testing. AI systems will trade off performance and available network or computing resources in these real environments. For multimedia research, as we use HAR to power interactive multimedia systems, we reflected this in our design considerations. For example, our method is inherently robust against packet loss as it requires collecting data in a broad sampling window. However, the window causes a delay in detecting when a new activity has started. This can be undesirable in health tracking where short interval training needs detection; for everyday tracking, as is our long-life use case, such a delay is less of an issue.

6 CONCLUSION & FUTURE WORK

In this article, we have presented a system and method using a *Convolutional Neural Network* for multi-class, real-time *Human Activity Recognition* using a low cost/low power BLE sensor. We demonstrate 94.79% accuracy on 6 activity classes using the SmartLab-v1 dataset. While this is approximately 2% lower than other state of the art techniques, we do so using an order of magnitude less data at 1Hz, using only using a tri-axial accelerometer, and estimate 1 year battery life using a single coin cell battery powered sensor. We then applied our method for use in a wearable fashion object for use in a smart night club environment. Our primary limitation is that the HAR is done using a nearby device. If that device is also portable (like a smartphone) then the battery of that device will incur in an extra load. That said, people are used to charging a phone regularly and many current phones and operating systems have optimizations for machine learning.

Embedding a lightweight long battery life sensor for HAR is ideal for deployments of smart materials and wearables. In particular, sensing with little data and without a connection enables us to perform studies. We illustrated a deployment where the battery needed to last weeks as a proof-of-concept. From here we look at performing long-term studies with wearable prototypes. Instead of building a wearable for a day of testing, we are now building prototypes for months of real use and testing. We believe this will not only allow us to collect more quantitative data, but will also let us perform longitudinal research that studies how people use and interact with smart textiles and wearables.

ACKNOWLEDGMENTS

This work was supported in part by RedBull. The authors would like to additionally thank the ByBorre fashion studio for their collaboration with designing the wearable bracelets. Additionally, the authors thank Judith Redi and Dick Bulterman for their technical advisement.

REFERENCES

[1] Mohammad Abu Alsheikh, Ahmed Selim, Dusit Niyato, Linda Doyle, Shaowei Lin, and Hwee-Pink Tan. 2016. Deep activity recognition models with triaxial accelerometers. In *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI, Phoenix, AZ, USA, 7.

[2] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz. 2013. A Public Domain Dataset for Human Activity Recognition using Smartphones.. In *Proceedings of the 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. ESANN, Bruges, Belgium, 437–442.

[3] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns Between Wellbeing Data and Context in Natural Language to Promote Behavior Change. *ACM Trans. Comput.-Hum. Interact.* 20, 5, Article 30 (Nov. 2013), 27 pages. https://doi.org/10.1145/2503823

[4] Frank R. Bentley and Crysta J. Metcalf. 2007. Sharing Motion Information with Close Family and Friends. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. ACM, New York, NY, USA, 1361–1370. https://doi.org/10.1145/1240624.1240831

[5] Sergio Cabrero, Jack Jansen, Thomas Röggla, John Alexis Guerra-Gomez, David A. Shamma, and Pablo Cesar. 2017. CWI-ADE2016 Dataset: Sensing Nightclubs through 40 Million BLE Packets. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (Taipei, Taiwan) *(MMSys'17)*. Association for Computing Machinery, New York, NY, USA, 181–186. https://doi.org/10.1145/3083187.3083213

[6] Pierluigi Casale, Oriol Pujol, and Petia Radeva. 2011. Human activity recognition from accelerometer data using a wearable device. In *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, Las Palmas de Gran Canaria, Spain, 289–296.

[7] Zhide Chen, Jiyun Wu, Aniello Castiglione, and Wei Wu. 2016. Human continuous activity recognition based on energy-efficient schemes considering cloud security technology. *Security and Communication Networks* 9, 16 (2016), 3585–3601.

[8] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding Quantified-selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 1143–1152. https://doi.org/10.1145/2556288.2557372

[9] Sunny Consolvo, Predrag Klasnja, David W. McDonald, Daniel Avrahami, Jon Froehlich, Louis LeGrand, Ryan Libby, Keith Mosher, and James A. Landay. 2008. Flowers or a Robot Army?: Encouraging Awareness & Activity with Personal, Mobile Displays. In *Proceedings of the 10th International Conference on Ubiquitous Computing* (Seoul, Korea) *(UbiComp '08)*. ACM, New York, NY, USA, 54–63. https://doi.org/10.1145/1409635.1409644

[10] Juan Davila, Ana-Maria Cretu, and Marek Zaremba. 2017. Wearable Sensor Data Classification for Human Activity Recognition Based on an Iterative Learning Framework. *Sensors* 17, 6 (Jun 2017), 1287. https://doi.org/10.3390/s17061287

[11] Nicholas Diliberti, Chao Peng, Christopher Kaufman, Yangzi Dong, and Jeffrey T. Hansberger. 2019. Real-Time Gesture Recognition Using 3D Sensory Data and a Light Convolutional Neural Network. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 401–410. https://doi.org/10.1145/3343031.3350958

[12] Daniel A. Epstein, Alan Borning, and James Fogarty. 2013. Fine-grained Sharing of Sensed Physical Activity: A Value Sensitive Approach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. ACM, New York, NY, USA, 489–498. https://doi.org/10.1145/2493432.2493433

[13] Wenchao Jiang and Zhaozheng Yin. 2015. Human Activity Recognition using Wearable Sensors by Deep Convolutional Neural Networks. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, Brisbain, Australia, 1307–1310.

[14] Wenchao Jiang and Zhaozheng Yin. 2015. Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks. In *Proceedings of the 23rd ACM International Conference on Multimedia* (Brisbane, Australia) *(MM '15)*. Association for Computing Machinery, New York, NY, USA, 1307–1310. https://doi.org/10.1145/2733373.2806333

[15] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler. 2006. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Transactions on Information Technology in Biomedicine* 10, 1 (Jan 2006), 156–167. https://doi.org/10.1109/TITB.2005.856864

[16] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2015), 15.

[17] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12, 2 (2011), 74–82.

[18] Oscar D Lara and Miguel A Labrador. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials* 15, 3 (2013), 1192–1209.

[19] Jin Lee and Jungsun Kim. 2016. Energy-efficient real-time human activity recognition on smart mobile devices. *Mobile Information Systems* 2016 (2016), 12.

[20] Yunji Liang, Xingshe Zhou, Zhiwen Yu, and Bin Guo. 2014. Energy-efficient motion related activity recognition on mobile devices for pervasive healthcare. *Mobile Networks and Applications* 19, 3 (2014), 303–317.

[21] Jeffrey W Lockhart, Gary M Weiss, Jack C Xue, Shaun T Gallagher, Andrew B Grosner, and Tony T Pulickal. 2011. Design considerations for the WISDM smart phone-based sensor mining architecture. In *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*. ACM, San Diego, CA, USA, 25–33.

[22] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 16, 1 (2016), 115.

[23] Alfredo J Perez, Miguel A Labrador, and Sean J Barbeau. 2010. G-sense: a scalable architecture for global sensing and monitoring. *IEEE Network* 24, 4 (2010), 57–64.

[24] R. W. Picard. 2016. Automating the Recognition of Stress and Emotion: From Lab to Real-World Impact. *IEEE MultiMedia* 23, 3 (July 2016), 3–7. https://doi.org/10.1109/MMUL.2016.38

[25] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. 2014. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, Colombus, Ohio, USA, 512–519. https://doi.org/10.1109/CVPRW.2014.131

[26] Jorge Luis Reyes-Ortiz, Luca Oneto, Albert Samà, Xavier Parra, and Davide Anguita. 2016. Transition-Aware Human Activity Recognition Using Smartphones. *Neurocomputing* 171 (2016), 754–767.

[27] Charissa Ann Ronao and Sung-Bae Cho. 2016. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications* 59 (2016), 235–244.

[28] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. 2015. A survey of online activity recognition using mobile phones. *Sensors* 15, 1 (2015), 2059–2085.

[29] Salwa Slim, Ayman Atia, Marwa M.A., and Mostafa-Sami Mostafa. 2019. Survey on Human Activity Recognition based on Acceleration Data. *International Journal of Advanced Computer Science and Applications* 10 (01 2019), 84–98. https://doi.org/10.14569/IJACSA.2019.0100311

[30] Niall Twomey, Tom Diethe, Xenofon Fafoutis, Atis Elsts, Ryan McConville, Peter Flach, and Ian Craddock. 2018. A Comprehensive Study of Activity Recognition Using Accelerometers. *Informatics* 5, 2 (2018), 37. https://doi.org/10.3390/informatics5020027

[31] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. 2019. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters* 119 (2019), 3 – 11. https://doi.org/10.1016/j.patrec.2018.02.010 Deep Learning for Pattern Recognition.

[32] Zhixian Yan, Vigneshwaran Subbaraju, Dipanjan Chakraborty, Archan Misra, and Karl Aberer. 2012. Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach. In *2012 16th international symposium on wearable computers*. IEEE, Cancún, Mexico, 17–24.

[33] Jian Bo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. 2015. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, Buenos Aires, Argentina, 25–31.

[34] Jie Yin, Qiang Yang, and Jeffrey Junfeng Pan. 2008. Sensor-based abnormal human-activity detection. *IEEE Transactions on Knowledge and Data Engineering* 20, 8 (2008), 1082–1090.

[35] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. 2014. Convolutional neural networks for human activity recognition using mobile sensors. In *Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on*. EAI, Austin, Texas, USA, 197–205.

[36] Mi Zhang and Alexander A. Sawchuk. 2012. USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. ACM, New York, NY, USA, 1036–1043. https://doi.org/10.1145/2370216.2370438