Adaptive Mobile Sensing: Leveraging Machine Learning for Efficient Human Behavior Modeling

A Technical Report submitted to the Department of Systems and Information Engineering

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Ian Tucker
Spring, 2020

Technical Project Team Members
Erin K. Barrett
Cameron M. Fard
Hannah N. Katinas
Charles V. Moens
Lauren E. Perry
Blake E. Ruddy
Shalin D. Shah
Tucker J. Wilson

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Adaptive Mobile Sensing: Leveraging Machine Learning for Efficient Human Behavior Modeling

Erin K. Barrett, Cameron M. Fard, Hannah N. Katinas, Charles V. Moens, Lauren E. Perry, Blake E. Ruddy, Shalin D. Shah, Ian S. Tucker, Tucker J. Wilson, Mark Rucker, Lihua Cai, Laura E. Barnes, Mehdi Boukhechba
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, Virginia, United States
ekb6ee, cmf6gh, hnk8er, cvm8aj, lep2jb, ber5rk, sds3bd, ist5te, tjw2kt, mr2an, lc3cp, lb3dp, mob3f@virginia.edu

Abstract — **Smartphones can collect millions of data points from each of its users daily, contributing to a significant change in how the healthcare community approaches health monitoring. This paper provides a framework for how smartphone sensor data can be collected, cleaned, stored, and modeled to effectively predict human states as a step towards health monitoring. To develop robust contextual models, a three-week study was conducted to collect data through a mobile crowdsensing application named Sensus. In this study, participants used multiple sensing strategies, ranging from infrequent sampling to continuous sampling, to determine the effect each has on data integrity and battery life. For a future study, a dynamic data collection strategy was developed that uses a machine learning model trained on existing data collected from 220 participants to forecast when a smartphone will be active and trigger sensor sampling accordingly. Results of this study include 1) extraction of model features that deliver maximized data quality with minimized battery consumption as compared to pre-existing baseline models, 2) implementation of context-driven modeling of user smartphone data on user's contextual environment, and 3) customization of a time-series database for optimized data queries used in metadata visualizations. The adaptive sensing models produced could be used in future large population studies that efficiently examine patterns of behavior in multiple individuals over extended periods to identify disease indicators present in an average user's daily life.**

***Keywords* — mobile sensing, adaptive sensing, machine learning**

INTRODUCTION

Development of mobile sensing technologies in personal devices, such as smartphones and wearable technology, has the potential to understand human behaviors and their embedded context. One implication of these advancements would enable users to detect a disease earlier than traditional health monitoring methods, resulting in faster treatment of symptoms while preventing the spread of illness. In order to develop models to predict the presence of diseases, smartphones need to efficiently collect millions of data points through mobile sensing. This study investigated three mobile sensing strategies to determine a balance between data quality and battery usage.

Continuous, high-sampling smartphone data collection would maximize data utility, but could severely increase smartphone battery consumption. Sparse, intermittent smartphone data collection would likely lead to very limited predictive performance. This brings about the need for a sensing strategy that adapts data collection based on smartphone sensor activity.

In this study, a crowdsensing application was used to collect mobile sensing data and administer context surveys in three different sensing strategy protocols: background, foreground, and static adaptive sensing. Predictive models from the newly collected data sets were developed. Additionally, a dynamic adaptive sensing strategy was built. Future data collection could provide a larger sample set which could further enhance predictive models, and analysis of updated models could provide further insight on the implications of this research.

RELATED WORK

In recent years, smartphone use has drastically increased as technology becomes more integrated into Americans' lives. A Pew Research Center (2019) survey found that 81% of Americans own smartphones [1]. These devices are capable of collecting millions of data points through sensors such as GPS, accelerometer, gyroscope, microphone, camera and Bluetooth [2]. This has led to an increase in the popularity of health-related applications that use these sensors to diagnose and predict health outcomes. About 58% of smartphone users have downloaded a health-related application [3]. Many studies have been dedicated to recognizing user context and activities to understand the user's status, which can ultimately be used to predict their health outcomes. One study used a system of wearables to collect sensor data and extract features to classify human activities with 94% accuracy [4].

While many mobile phone apps can connect to wearables to track and predict health outcomes, recent studies have focused on using smartphone sensors instead as a passive way to collect data. One study used raw GPS data to determine users' transportation mode, such as walking and driving [5]. Another study used solely accelerometer data to develop a predictive model to recognize ambulatory activities such as walking, jogging, climbing stairs, sitting and standing [6]. Additionally, Dernbach et. al. analyzed smartphone sensor data to recognize complex user activities not specifically tied to explicit data points, such as cooking and cleaning. While the researchers found that complex activities are much harder to recognize than simple activities, the performance recognizing complex activities had an accuracy of over 50% [7]. Otebolaku and Andrade found that incorporating additional data points such as orientation and rotation in their study of accelerometer data improved the performance of classifying users' context [8]. Additional studies have investigated different mobile sensing strategies in order to efficiently collect the data while minimizing the use of the battery life. This includes a study conducted with a static adaptive mobile sensing strategy and a duty cycling mobile sensing strategy [9,10].

Some studies use these sensors to obtain health-related results by using different algorithms and classification models to understand a user's activity. For example, one paper proposed a smartphone application that would be able to detect sound-related respiratory symptoms (sneezes, coughs, sniffles, and throat clearing) that could occur in a user's everyday life [11]. Mental health monitoring and prediction from smartphone sensors has also been explored [12, 13]. Such studies have shown the vast benefits of mobile health applications that use these context detection methods. These benefits include early detection, reduced healthcare costs, and no additional purchases of devices since these applications can run smoothly on users' smartphones [14, 15]. This paper builds upon previous mobile sensing studies to investigate the impact of sensing strategies on data quality and utility at predicting users' context.

Methodology

I. Data Collection

Data were collected using the Sensus crowdsensing application, a mobile sensing system available on both iOS and Android [16]. Sensus can collect data from multiple smartphone sensors, including accelerometer, altitude, attitude, battery, compass, gyroscope, image metadata, and location. Sensus allows for customization of sensing protocols such that sensor activity and sampling frequency can be configured.

Three separate protocols were used during the study. Seven participants ran each protocol for one week each. The first protocol employed a static adaptive sensing (SAS) strategy with the accelerometer probe. Fig. I illustrates how a general adaptive sensing strategy operates. In the static adaptive sensing strategy, the active observation duration was 10 seconds, and the action interval (idle period) between active observations was 5 minutes. If the acceleration in the horizontal direction was greater than 0.2 g or less than -0.2 g, continuous, high-rate sensing was triggered for all enabled sensors. These time periods and thresholds were held static.

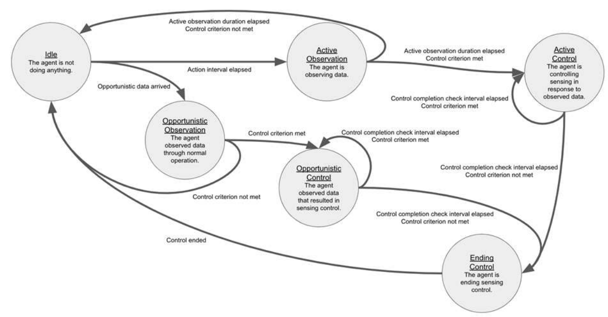

Figure I. Adaptive sensing flow [16]

Participants were required to keep Sensus open in the smartphone's background (i.e. the application could not be "swiped" out of) for the protocol to passively operate. An example segment of the process is detailed by Fig. II.

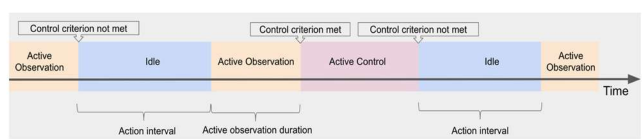

Figure II. Possible sequence of adaptive sensing states [16]

The second protocol only collected data when Sensus was opened (i.e. app in the foreground). No data were sampled passively when the application was in the background. Continuous, high-rate sensing was triggered for all enabled sensors when Sensus was open in the foreground.

The third protocol employed a GPS sensor listening strategy. By continuously listening to changes in GPS coordinates, Sensus was able to stay alive in the background and collect continuous data for all enabled sensors. Participants were required to keep Sensus open in the smartphone's background for the protocol to passively operate.

Each of the aforementioned protocols included a survey element via the Sensus app. The surveys served two purposes in the study: 1) collecting context information to allow tracking of the quality of data, and 2) collecting ground truth labels for users' activities to allow building context recognition algorithms using smartphone sensor data. The surveys collected data on the user's previous activity, the activity length, phone position, and more. Further information on the three surveys is included in Table I. Users were notified by a push notification when a survey arrived based on its scheduled time. Screens of an example survey are displayed in Fig. III.

TABLE I. SURVEY DESCRIPTIONS

Survey Set	Time Pushed	Question Types
Daily Random	Random, 8:00 AM - 8:00 PM	Prior phone position, prior body position, prior activity
Daily Morning	8:00 AM	Sleep duration, sleep quality, prior phone position, prior body position, prior activity
Daily Evening	8:00 PM	Phone interaction, prior phone position, prior body position, prior activity

II. Data Storage and Cleaning

After data collection, data preprocessing was conducted in order to utilize them for metadata queries, and real-time visualizations. The collected raw data were originally stored in zipped JSON form in an AWS S3 bucket. The queries and visualizations to be used would be primarily involving time as a variable, so it was determined that a time series database was the best approach for storing and querying the data. After researching different time series databases, InfluxDB was chosen over a database built on a traditional SQL architecture due to its increased performance in insertion, selection, and metadata queries. At this point, the rest of the process was developed as shown in Fig. IV. As a full set of data is collected, this process can be used by future researchers to store and visualize data.

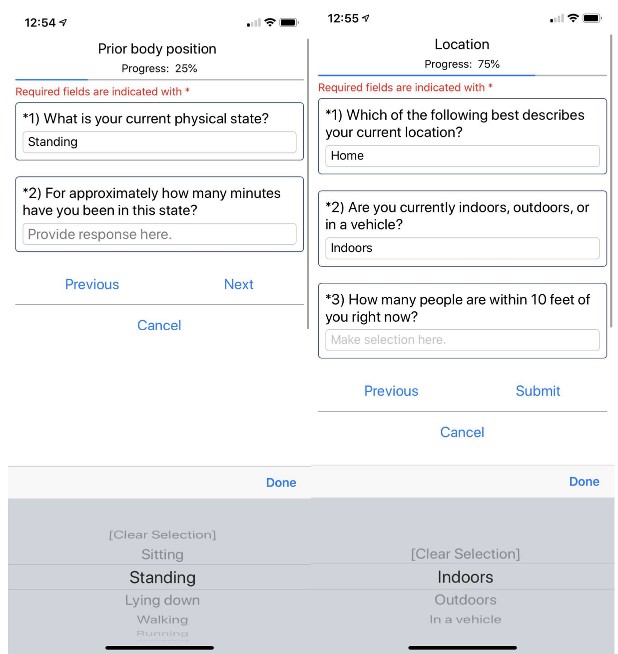

Figure III. "Prior body position" and "Location" survey screens

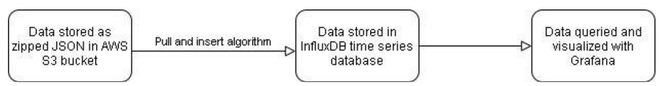

Figure IV. Data storage process

RESULTS

I. Comparing Data Quantity and Battery Performance

Quantity of data collected by each sensor varied between each sensing strategy. The number of data points collected per hour on average across all seven participants is summarized in Fig. V for the following sensors: accelerometer, altitude (not foreground protocol), attitude, compass (not foreground protocol), gyroscope, and location (not foreground protocol). As anticipated, the GPS listening protocol collected the most data on average, indicating Sensus was active more frequently in the background for the protocol. The foreground protocol collected significantly less data on average, as Sensus was not active in the background. Finally, the static adaptive sensing protocol fell in between the background and foreground strategies.

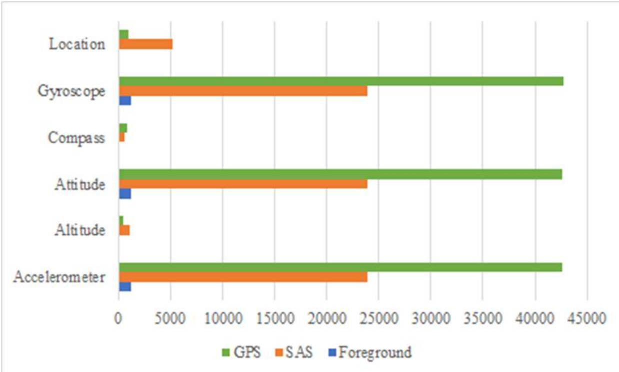

Figure V. Average data counts per hour by protocol

Sensus collects smartphone power connect plug-in and unplug events if the app is active. To assess the relative effect the protocols have on battery life, average time between plug-in events and average time plugged in were calculated and compared. Because it was expected that the foreground sensing strategy is only active for short periods of time, that protocol was excluded from the comparison. Periods of time less than 15 minutes and greater than 1,440 minutes (24 hours) were removed.

Time between power (charging) plug-in events was very similar between the GPS listening protocol and static adaptive sensing protocol on average. Time plugged in was 11% longer for the GPS listening protocol versus the static adaptive sensing protocol. Fig. VI summarizes these results.

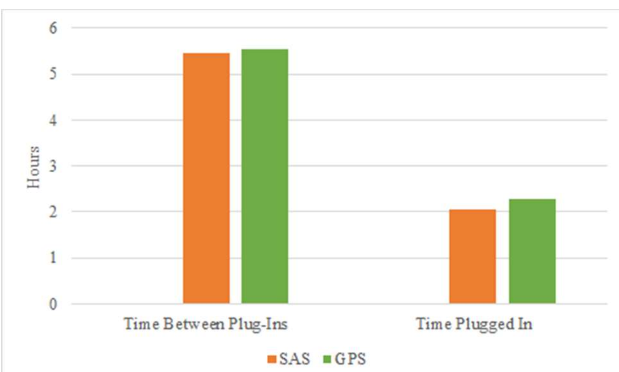

Figure VI. Time between distinct plug-in events and time plugged in by protocol

II. Comparing the Impact of Sensing Strategies on the Performance of Context Detection Models

Impact on data utility was assessed by producing and testing predictive models for four contextual features: phone location, user location, user activity, and user physical state. These features, their subsequent classes, as well as frequency within each collection method are shown in Figure VII.

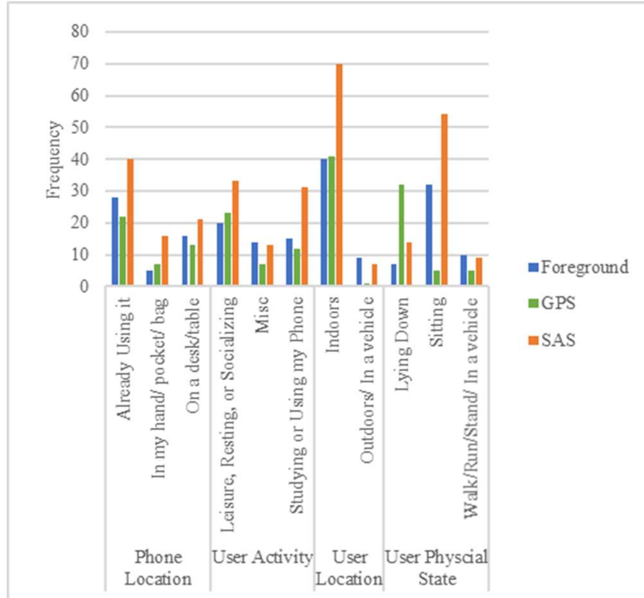

Figure VII. Frequency of features by class for each protocol

Phone location describes where the phone was in relation to the user at the time of the survey. User activity describes what the user was doing before the survey. User location describes where the user was before the survey. User physical state describes what the user was physically doing before the survey. Predictors used on these features are as follows, each recorded at time of survey response:

- Magnitude of linear acceleration in X, Y, and Z directions
- Magnitude of gyroscope in X, Y, and Z directions
- Latitude & longitude
- Hour of the day
- Day of the week

Random forest was used to create models for each feature and cross-validation was used to test model performance in predicting the contextual labels. This was done for each collection method, with F1 scoring used as metrics to compare effects on model performance as a result of the different data collection methods. Hypothetically, contextual detection models made from data from the GPS protocol should yield the highest F1 score as compared to models produced by data from the other two collection methods. The desire was for data from the SAS protocol to be able to produce predictive models whose F1 score was similar or equal to that of the models produced from GPS protocol data. This would mean the utility of the data collected by the SAS protocol was similar to that produced by the GPS protocol.

Leave one out cross-validation was used to assess the ability to produce quality predictive models depending on the data collection method. Results are shown in Table II for each collection method and feature/predictand across both metrics.

TABLE II. F1 SCORES

Feature	SAS	Foreground	GPS
Phone Location	0.648	0.590	0.605
User Location	0.896	0.838	-
User Activity	0.625	0.594	0.618
User Physical State	0.378	0.353	0.618

Results from the F1 scores suggest that data from SAS collection can generate detection models which outperform any produced by foreground or GPS collection, save for the user's physical state feature. This suggests that data utility from the SAS protocol is equal to that of data collected through the GPS protocol.

The largest limitation for context detection model assessment regards the data collection periods being run in series rather than parallel in time, leading to varying distribution of classes for each data collection method. This is especially apparent in GPS, where the collection period overlapped with the 2020 Coronavirus Pandemic, and the subsequent quarantine orders reduced the variety of classes within each feature.

III. Investigating a Dynamic Adaptive Sensing Strategy

Using smartphone data previously collected from 220 users running Sensus for two weeks each, a dynamic adaptive sensing model that predicts when a user’s phone is in use was investigated. The ultimate aim of this model is to reduce Sensus’ battery consumption by turning listeners on smartphone sensors on/off based on whether the phone is currently in use [17]. The models were generated as follows: 1) the data were segmented into repeating intervals for prediction, 2) the data were transformed from time series accelerometer data into machine learning readable feature,; and 3) both individual models for each user’s data and a global model using all user data were trained and evaluated. In practice, this model would be running constantly, collecting accelerometer data at short intervals and then making predictions based on those intervals. This process was simulated using previously collected data.

The original data consisted of linear (corrected for gravity) acceleration on the x-, y- and z-axes, collected from a smartphone accelerometer at a rate of once per second. Note that the data transformation and model generation processes were the same for the individual and global models. For the global model, all user data were combined into a single data source, and the global model was trained and tested on that data.

That data were first transformed into a single 3-dimensional acceleration column and then segmented into repeating intervals for testing. Those intervals were then segmented into two sub-windows: a listening window and a prediction window. The listening window is the small time slice at the beginning of the interval where the smartphone accelerometer would be switched on, while the prediction window is the rest of the interval for which the state of the phone would be predicted. The algorithm transforms data in the listening window into a single row of features and the data in the prediction window into a single value representing the phone’s state: active/in-use and inactive/not in-use. The features generated are shown in Table III.

TABLE III. FEATURES OF TIME SERIES MACHINE LEARNING MODEL

Name	Description
Hour	Hour of day (in 24-hour time)
Day of Week	Day of Week
Weekend	True/false based on whether interval was recorded on a Friday/Saturday
Mean	Average acceleration in listening window
Median	Median acceleration in listening window
Standard Deviation	Standard deviation of acceleration in listening window
Range	Range of acceleration in listening window
Mean Lag 1, 2, and 3	Mean value from the first, second, and third previous listening windows

The phone’s state during the prediction window was determined using mean acceleration in the prediction window. Based on live gathered smartphone accelerometer data, the median 3-dimensional linear acceleration of a phone resting on a flat, stationary surface is 0.1 m/s. Therefore, if the mean acceleration in the prediction window is greater than 0.1, then the phone’s state is active.

After testing interval lengths of 1, 2 and 5 minutes and listening window lengths of 5, 10, 20 and 30 seconds, an interval of 5 minutes with a listening sub-window of 20 seconds was chosen for the model. The final data consisted of rows of features extracted from 20 second listening windows as described above with a column of state values (0 or 1) representing the state for the 4 minutes and 40 seconds after the listening windows. Thus, each row of transformed data represented five minutes of time-series data.

Once this data were transformed, logistic regression models were trained both on data from individual users, and on the global data from all users. These models underwent hyperparameter tuning using a grid-search strategy. The models were then evaluated using k-fold cross-validation with the optimal hyperparameters. The model coefficients, and the model itself were all stored after the model was trained and evaluated. The evaluation results for the global logistic regression model are shown in Table IV..

TABLE IV. GLOBAL AND INDIVIDUAL LOGISTIC REGRESSION RESULTS

Metric	Global	Individual
Accuracy	89.6%	89.1%
Precision	88.5%	78.9%
Recall	76.6%	69.4%
F1 Score	0.821	0.737

Employing a logistic regression algorithm, the model has an F1 score of 0.821 which suggests the validity and usefulness of this model. The model has an accuracy level of 89.6%, a precision of 88.5%, and a recall of 76.6%. The model used the following parameters: {'C': 5, 'max_iter': 100, 'tol': 0.0001}.

This logistic regression was also compared against two other model algorithms: XGBoost and Random Forest. Both models were built using data collected from the 220 users as well. The XGBoost model has an F1 score of 0.835 which suggests this model is valid and useful. Additionally, the XGBoost model has an accuracy level of 86.3%, a precision of 85.5%, and recall of 81.6%. The model used the following parameters: {'gamma': 0.4, 'learning_rate': 0.05, 'min_child_weight': 1, 'colsample_bytree': 0.4, 'max_depth': 10}. The Random Forest model achieved an F1 score of 0.833. It has similar results to the results of the XGBoost model: an accuracy level of 86.2%, a precision of 85.4%, and recall of 81.3%.

CONCLUSION

This study presents an assessment of the relative efficiency and utility of multiple smartphone sensing strategies' and proposes a dynamic adaptive sensing strategy to effectively balance efficiency and utility. Implementation and further development of the suggested dynamic adaptive sensing strategy can help better detect smartphone users' states and actions while reasonably preserving device battery life. Future applications could empower more significant smartphone integration in healthcare systems.

The amount and quality of predictors for context detection models depends primarily, but not solely, on data quantity. Data quantity from the GPS listening protocol and static adaptive sensing protocol far exceeded that of the foreground protocol, and the former two protocols outperformed the latter protocol in context detection. Notably, the data from static adaptive sensing strategy generated better performing models for the phone location, user location, and user activity features than data from the GPS listening sensing strategy, which outperformed for user physical state. This indicates the importance of timely data collection, which is a primary purpose of adaptive sensing. Though battery consumption was very similar between the two protocols, the large quantity of data generated from the GPS listening protocol suggests that smartphone battery life would be more adversely affected over time.

A limitation of the comparison of data from the protocols is that the study spanned a period prior to and during a heightened period of the 2020 Coronavirus Pandemic. Another limitation is that because of the small participant pool, numerous survey response instances were not large enough to create context detection models for. Furthermore, certain sensors were active for some protocols or one protocol but not for others. Future studies should make sensors "turned on" uniform across protocols and enlist larger participant pools with more consistent environments.

The recommended dynamic adaptive sensing strategy performed well on all metrics and can be implemented in Sensus for data collection and analysis. Additional research could improve clarity in the efficiency and utility of multiple dynamic adaptive sensing strategies that use different and/or additional features, prediction intervals, and other input variables. Ultimate deployment of an optimized dynamic adaptive sensing protocol could allow for the generation of highly accurate and personalized context detection models, the widespread adoption of which would transform predictive health.

ACKNOWLEDGMENT

We give our sincere thanks to our faculty advisors, Department of Engineering Systems and Environment, and the University of Virginia School of Engineering and Applied Science for their support. Our research was funded in part by the Defense Advanced Research Projects Agency (DARPA) Warfighter Analytics using Smartphones for Health (WASH) Program.

REFERENCES

[1] "Mobile Fact Sheet." Demographics of Mobile Device Ownership and Adoption in the United States. *Pew Research Center: Internet, Science & Tech*.

[2] Laurila, J., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T.-M.-T., Dousse, O., Eberle, J., Miettinen, M. (2012). *The mobile data challenge: Big data for mobile computing research*. Nokia Research Center.

[3] Krebs, P., & Duncan, D. T. (2015). Health app use among US mobile phone owners: A National survey. JMIR MHealth and UHealth, 3(4), e101.

[4] Casale P., Pujol O., Radeva P. (2011) Human Activity Recognition from Accelerometer Data Using a Wearable Device. In: Vitrià J., Sanches J.M., Hernández M. (eds) Pattern Recognition and Image Analysis Lecture Notes in Computer Science

[5] Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008). Learning transportation mode from raw gps data for geographic applications on the web. *Proceedings of the 17th International Conference on World Wide Web - WWW '08*, 247.

[6] Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, *12*(2), 74–82.

[7] Dernbach, S., Das, B., Krishnan, N. C., Thomas, B. L., & Cook, D. J. (2012). Simple and Complex Activity Recognition through Smart Phones. *2012 Eighth International Conference on Intelligent Environments*, 214–221.
[8] Otebolaku, A. M., & Andrade, M. T. (2016). User context recognition using smartphone sensors and classification models. *Journal of Network and Computer Applications*, *66*, 33–51.
[9] L. Cai, M. Boukhechba, N. Kaur, C. Wu, L. E. Barnes and M. S. Gerber, "Adaptive Passive Mobile Sensing Using Reinforcement Learning," *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, Washington, DC, USA, 2019, pp. 1-6.
[10] Wang, Y.; Lin, J.; Annavaram, M.; Jacobson, Q.A.; Hong, J.; Krishnamachari, B.; Sadeh, N. A framework of energy efficient mobile sensing for automatic user state recognition. Proceedings of the 7th international conference on mobile systems, applications, and services. ACM, 2009, pp. 179–192.
[11] Xiao Sun, Zongqing Lu, Wenjie Hu, and Guohong Cao. (2015). SymDetector: detecting sound-related respiratory symptoms using smartphones. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15). Association for Computing Machinery, 97–108.
[12] Boukhechba, M., Daros, A.R., Fua, K., Chow, P.I., Teachman, B.A., & Barnes, L.E. (2018). DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. *Smart Health, 9-10,* 192-203.
[13] Boukhechba, M., Chow, P., Fua, K., Teachman, B.A., & Barnes, L.E. (2018). Predicting Social Anxiety From Global Positioning System Traces of College Students: Feasibility Study. *JMIR Mental Health, 5*(3), e10101.
[14] Agu, E., Pedersen, P., Strong, D., Tulu, B., He, Q., Wang, L., & Li, Y. (2013). "The smartphone as a medical device: Assessing enablers, benefits and challenges," *2013 IEEE International Conference on Sensing, Communications and Networking (SECON),* New Orleans, LA. 76–80.
[15] Otebolaku, A. M., & Andrade, M. T. (2016). User context recognition using smartphone sensors and classification models. *Journal of Network and Computer Applications*, *66*, 33–51.
[16] H. Xiong, Y. Huang, L. E. Barnes, and M. S. Gerber. Sensus: A Cross-Platform, General-Purpose System for Mobile Crowdsensing in Human-Subject Studies. *In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 415-426. ACM, 2016.
[17] Cai, L., Boukhechba, M., Kaur, N., Wu, C., Barnes, L.E., Gerber, M.S. (2019). Adaptive Passive Mobile Sensing Using Reinforcement Learning. *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks*.

AUTHOR INFORMATION

Erin K. Barrett, Cameron M. Fard, Hannah N. Katinas, Charles V. Moens, Lauren E. Perry, Blake E. Ruddy, Shalin D. Shah, Ian S. Tucker, Tucker J. Wilson, Undergraduate Students, Systems Engineering, University of Virginia.

Mark Rucker, Lihua Cai, Graduate Students, Systems Engineering, University of Virginia.

Laura E. Barnes, Ph.D., Associate Professor, Department of Engineering Systems and Environment, University of Virginia.

Mehdi Boukhechba, Ph.D., Assistant Professor, Department of Engineering Systems and Environment, University of Virginia.