Activity Recognition in Older Adults with Training Data from Younger Adults: Preliminary Results on in Vivo Smartwatch Sensor Data

Sabahat Fatima
sfatima@umd.edu
University of Maryland, College Park

ABSTRACT

Self-tracking using commodity wearables such as smartwatches can help older adults reduce sedentary behaviors and engage in physical activity. However, activity recognition applications that are typically deployed in these wearables tend to be trained on datasets that best represent younger adults. We explore how our activity recognition model, a hybrid of long short-term memory and convolutional layers, pre-trained on smartwatch data from younger adults, performs on older adult data. We report results on week-long data from two older adults collected in a preliminary study in the wild with ground-truth annotations based on activPAL, a thigh-worn sensor. We find that activity recognition for older adults remains challenging even when comparing our model's performance to state of the art deployed models such as the Google Activity Recognition API. More so, we show that models trained on younger adults tend to perform worse on older adults.

KEYWORDS

older adults, seniors, self tracking, human activity recognition, machine learning, smartwatches

ACM Reference Format:
Sabahat Fatima. 2021. Activity Recognition in Older Adults with Training Data from Younger Adults: Preliminary Results on in Vivo Smartwatch Sensor Data. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21), October 18–22, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3441852.3476475

1 INTRODUCTION

After a year of social distancing due to COVID-19, there is a concerning increase in sedentary behavior [33]. The lack of physical activity is a leading factor behind preventable chronic diseases such as diabetes, heart disease, and cancer, with older adults at a higher risk [5]. Prior work has shown that psychoeducational interventions can decrease sedentary behavior in older adults [25, 28] and technological interventions employing self-tracking can increase behavior awareness, yielding positive behavioral changes [10, 23, 43]. As we move away from deficit-based perspectives [16], there is a need for better activity recognition models to be employed in innovations that support older adults' agency, enjoyment, and overall well-being later in life [16].

Automated and less complex systems with low data capture burden are more effective and easily adoptable [8, 22]; hence smartwatches, which have evolved into health-tracking hubs, are promising candidates [31]. While studies show that older adults are interested in using self-tracking sensors [9], they have adopted these technologies less [13], as the current physical activity trackers do not effectively identify and track older adults' activities [40]. A reason behind this ineffectiveness can be that most self-tracking models are trained on younger adults whose activities and movement patterns differ from older adults [14]. A further problem is the scarcity of older adult data [19, 21, 30], with widespread human activity recognition (HAR) datasets representing younger adults between the ages of 18-48 [4, 24, 32].

In this study, we train an activity recognition model on younger adult smartwatch accelerometer data to explore how much the performance deviates when predicting younger versus older adults' (>65 years old) activities. We develop a hybrid model of long short-term memory and convolutional neural network layers (LSTM-CNN). Additionally, we compare our model to prior approaches both in academia and industry such as the multilayer perceptron [37] and the Google Activity Recognition API [1].

2 RELATED WORK

Through a systematic literature review, we find that there are three main smartwatch datasets, none of which represent older adults: WISDM [38, 39], UCI-HHAR [34], and the Extrasensory [36] dataset. In this work, we adopt the Extrasensory dataset, the largest, with 60 users (ages 18-42), 300K samples, and 51 activity labels. The samples are collected *in the wild*, or in a real-life setting, adding diversity and imbalance.

We observe that different activity recognition algorithms have been used with this dataset such as Logistic Regression [36], Multilayer Perceptron [37], CNN-Random Forest [11], Active Learning with Logistic Regression [3], and Boosting methods [35]. However, these studies typically include not only smartwatch data, as in this work, but also smartphone sensors with some of the highest accuracies reported being around 89%. More recent neural network architectures such as LSTM-CNN [42], have demonstrated higher performance on similar tasks that involve time series data [6, 7, 29, 41]. Thus, in this work we develop an LSTM-CNN model and train it on the Extrasensory dataset.

3 ACTIVITY RECOGNITION MODELS

Proposed Model. A hybrid LSTM-CNN model is proposed to extract spatio-temporal features from the data. The pre-processed

ASSETS '21, October 18–22, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8306-6/21/10.
https://doi.org/10.1145/3441852.3476475

data is first taken as input by a 2 layer LSTM with 64 and 32 neurons respectively to learn time-related features. Spatial features are next extracted using 2 convolutional layers of 128 filters each. A max-pooling layer is placed between the convolutional layers and the model ends with a global average pooling layer and a batch normalization layer. The output dense layer has a sigmoid activation layer to derive a probability distribution for classification.

Baseline Model. We attempt to replicate the Multilayer Perceptron (MLP) algorithm with two 16 neuron hidden layers used by Vaizman et al. [37], who developed the Extrasensory dataset. However, compared to their effort, we use only smartwatch accelerometer data, leaving out phone sensors and location insight for participant privacy.

The models were trained on the Extrasensory dataset to classify 5 primary participant activities: *Sitting, Walking, Running, In-Vehicle, and Standing*. These labels are selected as they are the only ones present in our older adult data that correspond to both the activPAL [2] and the Google Activity Recognition API labels. These labels are the primary labels to other fine-grained secondary labels classified by the model trained on the Extrasensory dataset like *Cooking* or *Computer Work*. Table 1 shows our label mapping method which will be used when calculating model accuracy. Furthermore, to address the challenge of unbalanced data, balanced accuracy (BA) is also measured.

4 PRELIMINARY STUDY AND DATA COLLECTION

Data Collection. We employ preliminary smartwatch accelerometer data from two older adults (PP1, PP2), collected as part of a larger project in a a week-long in vivo study. We obtained ground truth labels using activPAL [2], a thigh-worn activity tracker. We also obtain activity predictions from the participants' Google Wear OS smartwatch through the Google Activity Recognition API. To match the training Extrasensory dataset, the preliminary data was collected over a week using a 25Hz sampling frequency with 500 samples/minute in a 20-second window. To further match the training data's format, the older adult data was pre-processed to aggregate the same statistical features [36] per minute window.

Label Annotation. The annotations for our study were based on the the ground truth activPAL labels. However, there is no one-to-one mapping between the activity labels across activPAL, the Extrasensory dataset, and the Google API, used in this work. Thus, we estimate an alignment shown in Table 1. Some mappings are more straightforward such as (*Seated Transport*, *In-Vehicle*, *In-Vehicle*). Others are not, so we use hybrid labels. For example, (*Sedentary*, *Sitting*, *Tilting & Still or Still*) was chosen as the mapping that yielded the highest prediction across all plausible mappings.

5 EXPERIMENTS AND RESULTS

Performance on younger adult data. It is already challenging to classify *in the wild* activities for younger adults using multiple sensors, with Vaizman et. al. reporting an average 87% accuracy and 77% balanced accuracy (BA) across 25 labels [36]. Using a single sensor, performance is generally lower with Vaizman et. al. reporting an average 73% accuracy and 68% BA using only phone accelerometer data. On younger adult data from two participants (EP1, EP2) left out from the same dataset, our LSTM-CNN model has an average ~89% accuracy and ~67% BA across five labels. While the accuracy is ~2% greater than the multiple sensor model, the BA is 10% lower than the multiple sensor method and ~1% lower than the phone accelerometer model. This highlights the limitation of unbalanced data and the lesser amount of participants who provided watch data compared to phone data. Performance differed across labels, with LSTM-CNN working better on *Walking, Running,* and *In-Vehicle* labels. Lower performance is seen for *Standing* and *Sitting*, as the upper body could be engaged in diverse secondary activities across participants involving the wrist, while the lower body is still. Primary activities, like *Standing*, with diverse secondary activities have displayed lower accuracies in the literature even with mixed sensor models [36].

Performance on older vs. younger adult data. The model trained on younger adult data performs better on younger participants than on older participants. As seen in Fig. 1a,d[1], there is an average ~25% greater accuracy and ~11% greater BA across the selected activities for the younger participants compared to the older participants. For older adults, the accuracy and BA are much lower across all labels, except *Standing*, with a poor ~51% average BA. This agrees with previous literature on fitness tracking sensors inaccurately reporting older adult activities like Walking, especially at slower speeds [40]. While *Standing* had a higher accuracy for older adults (+9%), the BAs show very close numbers with a ~3% difference and EP2 having the greatest BA.

Past Models. While LSTM-CNN works better than the past MLP model for younger adults, for older adult participants, there are mixed results. LSTM-CNN performs better for *Walking* and *Running* while MLP performs better on *Sitting*. We also compared the Google API's performance with our model's performance on the older adults' data using activPAL data as ground truth as seen in Fig. 1c,f. In terms of accuracy and BA, our model performs better or comparably to the Google API for *Walking* (+4%,+1%), *Running* (+4%,+1%), and *Standing* (+46%,+12%). The Google API still performs better for *In-Vehicle* in terms of accuracy by 2.5% for PP1 and PP2 and BA for PP1 by 30%. For *Sitting*, the Google API performs better by 5% in terms of both BA for PP1 and PP2 and accuracy for PP2. These two activities depend less on lower body motion, which the ground truth (activPAL) tracker gains more insight from, than the other activities. This could give insight into the greater performance of the Google API on the two labels (*In-Vehicle*, *Sitting*). We consider the limits of assessing the industry model with ground truth from the activPAL which has limited labels gaining data from lower body motion. With that in mind, this shows that for two participants, the Google API is not optimal for older adult HAR.

Limitations. We present findings with two older adults and given that older adults' activities vary, the results are limited. Both the Extrasensory and our dataset have imbalanced data across the activities, which is common for data obtained in the wild (people tend to do some activities more than others). So for this analysis, we limited the types of activities in an attempt to constrain the imbalance. More importantly, there is a difference in how ground truth labels were obtained for younger vs. older adults (self-reported vs. activPAL), which could explain some of the differences.

[1] Color-blind safe and print friendly colors were chosen from: https://colorbrewer2.org.

Table 1: Extrasensory, activPAL, and Google Activity Recognition API Label alignment

activPAL Labels	Our Baseline Model Labels (Extrasensory Dataset)	Google API Labels
Sedentary	Sitting	Tilting & Still or Still
Upright & Sedentary	Standing	On Foot & Still or Still
Stepping (low cadence)	Walking	On Foot and Walking
Stepping (high cadence)	Running	On Foot and Running
Seated Transport	In-Vehicle	In-Vehicle

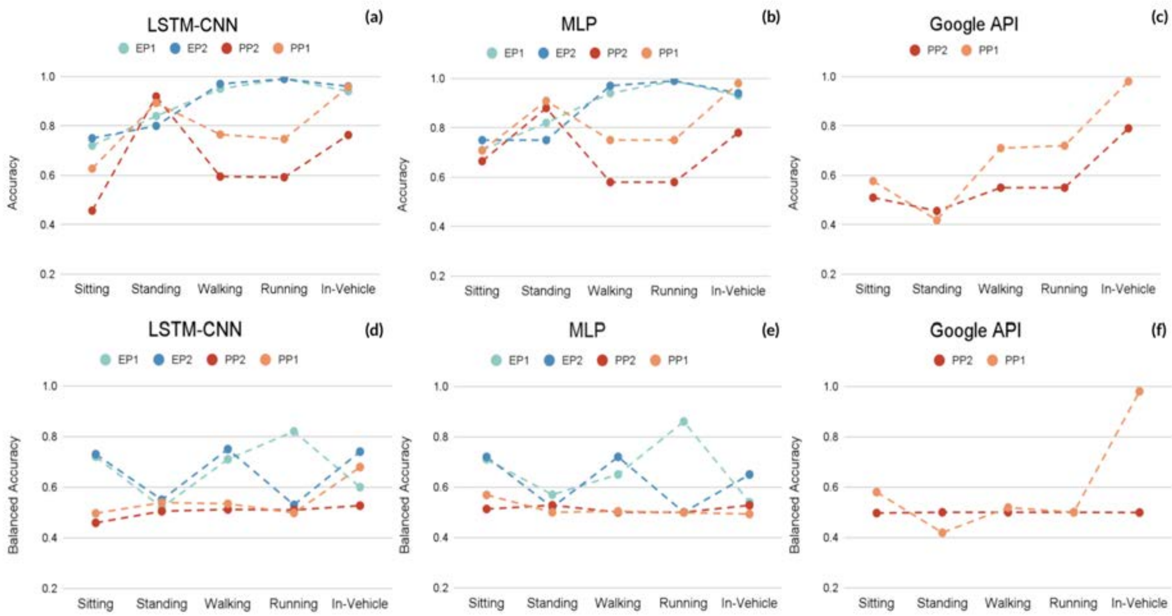

Figure 1: (a-c) Accuracy of younger adult participants (EP1, EP2) and older adults (PP1,PP2) using the LSTM-CNN, MLP, and Google API models respectively and activPAL ground truth data. (d-f) Balanced Accuracy (BA) of EP1, EP2, PP1, and PP2 using LSTM-CNN, MLP, and Google API models respectively and activPAL ground truth data.

6 CONCLUSION

After pre-training machine learning models on the largest younger adult smartwatch dataset (Extrasensory) and classifying 5 primary activities, our LSTM-CNN model performs comparably or better than past models like MLP on younger adult data and the Google API on older adult data. As hypothesized, the models trained on younger adult data generally perform less accurately on preliminary older adult data. While this represents two participants, the low BA of 51% shows a need for larger training datasets with older adult representation. The challenge of data scarcity, unbalanced data, and diverse self-reported secondary activities in both younger and older adults calls for model personalization through methods like teachable machines [12, 17, 18, 26], which leverage advances in transfer learning [20] and meta learning [15, 27] and empower end users to fine tune models to their idiosyncratic characteristics.

ACKNOWLEDGMENTS

I would like to thank the broader team that contributed the data as part of the larger project supported by NSF (#1955568): Young-Ho Kim, Diana Chou, Bongshin Lee, Margaret Danilovich, David Conroy, Amanda Lazar, Eun Kyoung Choe, and my advisor Hernisa Kacorri. I am grateful to our participants as well as Vaizman et al. [37] for sharing the Extrasensory dataset. I also appreciate feedback from fellow students (Rie Kamikubo and Lining Wang), and anonymous reviewers.

REFERENCES

[1] [n.d.]. Google Activity Recognition API. https://developers.google.com/location-context/activity-recognition/
[2] 2001. PAL Technologies. https://www.palt.com/
[3] Rebecca Adaimi and Edison Thomaz. 2019. Leveraging active learning and conditional mutual information to minimize data annotation in human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–23.
[4] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz, et al. 2013. A public domain dataset for human activity recognition using smartphones.. In *Esann*, Vol. 3. 3.

[5] Centers for Disease Control and Prevention. 2019. Lack of Physical Activity. https://www.cdc.gov/chronicdisease/resources/publications/factsheets/physical-activity.htm

[6] Rohan Chandra, Uttaran Bhattacharya, Christian Roncal, Aniket Bera, and Dinesh Manocha. 2019. RobustTP: End-to-End Trajectory Prediction for Heterogeneous Road-Agents in Dense Traffic with Noisy Sensor Inputs. In *ACM Computer Science in Cars Symposium* (Kaiserslautern, Germany) *(CSCS '19)*. Association for Computing Machinery, New York, NY, USA, Article 2, 9 pages. https://doi.org/10.1145/3359999.3360495

[7] Yuwen Chen, Kunhua Zhong, Ju Zhang, Qilong Sun, and Xueliang Zhao. 2016. LSTM networks for mobile human activity recognition. In *2016 International conference on artificial intelligence: technologies and applications*. Atlantis Press, 50–53.

[8] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A Epstein, Felicia Cordeiro, Matthew Kay, Gregory D Abowd, Tanzeem Choudhury, James Fogarty, et al. 2017. Semi-automated tracking: a balanced approach for self-monitoring applications. *IEEE Pervasive Computing* 16, 1 (2017), 74–84.

[9] Jane Chung, Blaine Reeder, Amanda Lazar, Jonathan Joe, George Demiris, and Hilaire J Thompson. 2014. Exploring an informed decision-making framework using in-home sensors: older adults' perceptions. *Journal of Innovation in Health Informatics* 21, 2 (2014), 73–77.

[10] David E Conroy, Jaclyn P Maher, Steriani Elavsky, Amanda L Hyde, and Shawna E Doerksen. 2013. Sedentary behavior as a daily process regulated by habits and intentions. *Health Psychology* 32, 11 (2013), 1149.

[11] Federico Cruciani, Anastasios Vafeiadis, Chris Nugent, Ian Cleland, Paul McCullagh, Konstantinos Votis, Dimitrios Giakoumis, Dimitrios Tzovaras, Liming Chen, and Raouf Hamzaoui. 2020. Feature learning for human activity recognition using convolutional neural networks. *CCF Transactions on Pervasive Computing and Interaction* 2, 1 (2020), 18–32.

[12] Utkarsh Dwivedi, Jaina Gandhi, Raj Parikh, Merijke Coenraad, Elizabeth Bonsignore, and Hernisa Kacorri. 2021. Exploring Machine Teaching with Children. In *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. Ieee.

[13] John Elflein. 2019. Percentage of people that used technology to track their fitness by age 2016. https://www.statista.com/statistics/742448/global-fitness-tracking-and-technology-by-age/

[14] José María Faílde Garrido, María Dolores Dapía Conde, Miguel Ángel Vázquez Vázquez, and Lucía Vázquez Rodríguez. 2021. Are Older Adults' Leisure Patterns Consistent with the Active Ageing Model? The Influence of the Ageist Stereotypes. *Social Policy and Society* (2021), 1–22. https://doi.org/10.1017/S1474746421000221

[15] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*. PMLR, 1126–1135.

[16] Kathrin Gerling, Mo Ray, Vero Vanden Abeele, and Adam B Evans. 2020. Critical reflections on technology to support physical activity among older adults: An exploration of leading HCI venues. *ACM Transactions on Accessible Computing (TACCESS)* 13, 1 (2020), 1–23.

[17] Jonggi Hong, Kyungjun Lee, June Xu, and Hernisa Kacorri. 2020. Crowdsourcing the Perception of Machine Teaching. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376428

[18] Hernisa Kacorri. 2017. Teachable machines for accessibility. *ACM SIGACCESS Accessibility and Computing* 119 (2017), 10–18.

[19] Hernisa Kacorri, Utkarsh Dwivedi, and Rie Kamikubo. 2020. Data Sharing in Wellness, Accessibility, and Aging. *NeurIPS 2020 Workshop on Dataset Curation and Security* (2020).

[20] Hernisa Kacorri, Kris M. Kitani, Jeffrey P. Bigham, and Chieko Asakawa. 2017. *People with Visual Impairment Training Personal Object Recognizers: Feasibility and Challenges*. Association for Computing Machinery, New York, NY, USA, 5839–5849. https://doi.org/10.1145/3025453.3025899

[21] Rie Kamikubo, Utkarsh Dwivedi, and Hernisa Kacorri. 2021. Sharing Practices for Datasets Related to Accessibility and Aging. *The23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21)* (2021). https://doi.org/10.1145/3441852.34712081

[22] Young-Ho Kim, Bongshin Lee, Arjun Srinivasan, and Eun Kyoung Choe. 2021. *Data@Hand: Fostering Visual Exploration of Personal Data On Smartphones Leveraging Speech and Touch Interaction*. Association for Computing Machinery, New York, NY, USA. https://doi-org.proxy-um.researchport.umd.edu/10.1145/3411764.3445421

[23] William J Korotitsch and Rosemery O Nelson-Gray. 1999. An overview of self-monitoring research in assessment and treatment. *Psychological Assessment* 11, 4 (1999), 415.

[24] 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.

[25] Matthew A Ladwig, Christopher N Sciamanna, Kayla Rutt, Joshua Blaker, Kalen Kearcher, Brandon J Auer, Liza S Rovniak, David E Conroy, Jinger S Gottschall, Matthew L Silvis, et al. 2021. Adult outdoor group sport play during a pandemic: Feasibility, acceptability, and program adherence results from a study of modifications to mitigate COVID-19 risk. *Preventive Medicine Reports* (2021), 101476.

[26] Kyungjun Lee, Jonggi Hong, Simone Pimento, Ebrima Jarjue, and Hernisa Kacorri. 2019. Revisiting Blind Photography in the Context of Teachable Object Recognizers. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 83–95. https://doi.org/10.1145/3308561.3353799

[27] Chenglin Li, Di Niu, Bei Jiang, Xiao Zuo, and Jianming Yang. 2021. Meta-HAR: Federated Representation Learning for Human Activity Recognition. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 912–922. https://doi.org/10.1145/3442381.3450006

[28] Jaclyn P Maher, Martin J Sliwinski, and David E Conroy. 2017. Feasibility and preliminary efficacy of an intervention to reduce older adults' sedentary behavior. *Translational behavioral medicine* 7, 1 (2017), 52–61.

[29] Ronald Mutegeki and Dong Seog Han. 2020. A CNN-LSTM Approach to Human Activity Recognition. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. 362–366. https://doi.org/10.1109/ICAIIC48513.2020.9065078

[30] Joon Sung Park, Michael S Bernstein, Robin N Brewer, Ece Kamar, and Meredith Ringel Morris. 2021. Understanding the Representation and Representativeness of Age in AI Data Sets. *arXiv preprint arXiv:2103.09058* (2021).

[31] Christy Pettey. 2018. Wearables Hold the Key to Connected Health Monitoring. https://www.gartner.com/smarterwithgartner/wearables-hold-the-key-to-connected-health-monitoring/

[32] Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Kilian Förster, Gerhard Tröster, Paul Lukowicz, David Bannach, Gerald Pirkl, Alois Ferscha, et al. 2010. Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh international conference on networked sensing systems (INSS)*. IEEE, 233–240.

[33] Hamilton Roschel, Guilherme G Artioli, and Bruno Gualano. 2020. Risk of increased physical inactivity during COVID-19 outbreak in older people: a call for actions.

[34] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. 2015. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*. 127–140.

[35] Pratik Tarafdar and Indranil Bose. 2021. Recognition of human activities for wellness management using a smartphone and a smartwatch: a boosting approach. *Decision Support Systems* 140 (2021), 113426.

[36] Yonatan Vaizman, Katherine Ellis, and Gert Lanckriet. 2017. Recognizing Detailed Human Context in the Wild from Smartphones and Smartwatches. *IEEE Pervasive Computing* 16, 4 (2017), 62–74. https://doi.org/10.1109/MPRV.2017.3971131

[37] Yonatan Vaizman, Nadir Weibel, and Gert Lanckriet. 2018. Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–22.

[38] G Weiss. 2019. WISDM smartphone and smartwatch activity and biometrics dataset data set.

[39] Gary M Weiss, Kenichi Yoneda, and Thaier Hayajneh. 2019. Smartphone and smartwatch-based biometrics using activities of daily living. *IEEE Access* 7 (2019), 133190–133202.

[40] Christopher K Wong, Helena M Mentis, and Ravi Kuber. 2018. The bit doesn't fit: Evaluation of a commercial activity-tracker at slower walking speeds. *Gait & posture* 59 (2018), 177–181.

[41] Kun Xia, Jianguang Huang, and Hanyu Wang. 2020. LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access* 8 (2020), 56855–56866. https://doi.org/10.1109/ACCESS.2020.2982225

[42] Zhenqi Xu, Shan Li, and Weihong Deng. 2015. Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. 141–145. https://doi.org/10.1109/ACPR.2015.7486482

[43] Chih-Hsiang Yang and David E Conroy. 2019. Feasibility of an outdoor mindful walking program for reducing negative affect in older adults. *Journal of aging and physical activity* 27, 1 (2019), 18–27.