HIERARCHICAL DEEP LEARNING MODEL WITH INERTIAL AND PHYSIOLOGICAL SENSORS FUSION FOR WEARABLE-BASED HUMAN ACTIVITY RECOGNITION

Dae Yon Hwang[§], *Pai Chet Ng*[§], *Yuanhao Yu*[†], *Yang Wang*[†],
Petros Spachos[⋆], *Dimitrios Hatzinakos*[§], *Konstantinos N. Plataniotis*[§]

[§]Electrical and Computer Engineering Dept., University of Toronto, ON, Canada
[†]Huawei Technologies Canada, ON, Canada
[⋆]School of Engineering, University of Guelph, ON, Canada

ABSTRACT

This paper presents a human activity recognition (HAR) system with wearable devices. While various approaches have been suggested for HAR, most of them focus on either 1) the inertial sensors to capture the physical movement or 2) subject-dependent evaluations that are less practical to real world cases. To this end, our work integrates sensing inputs from physiological sensors to compensate the limitation of inertial sensors in capturing the human activities with less physical movements. Physiological sensors can capture physiological responses reflecting human behaviors in executing daily activities. To simulate a realistic application, three different evaluation scenarios are considered, namely All-access, Cross-subject and Cross-activity. Lastly, we propose a Hierarchical Deep Learning (HDL) model, which improves the accuracy and stability of HAR, compared to conventional models. Our proposed HDL with fusion of inertial and physiological sensing inputs achieves 97.16%, 92.23%, 90.18% average accuracy in All-access, Cross-subject, Cross-activity scenarios, which confirms the effectiveness of our approach.

Index Terms**—** PPG, ACC, EDA, HAR, Deep Learning

1. INTRODUCTION

Nowadays, wearable devices embedded with heterogeneous sensors can recognize human behaviors outside the laboratory setting through health, emotion and activity monitoring. This paper focuses on human activity recognition (HAR) with wearable devices, which is one of the most eminent areas because of a great potentiality to be extended in many different domains, such as ambient assisted living and human-computer interaction [1]. Despite the availability of physiological sensors embedded in wearable devices, most of works in HAR only exploited the sensing input from inertial sensors, such as accelerometer (ACC), gyroscope, and magnetometer, to infer human activity. For example, [2] exploits the ACC signals from the wearable device to distinguish self-harming activities from other activities, whereas [3] extracts the characteristics of peaks detected from the ACC signals to predict human activity. There are also works considering the fusion of multiple sensing modalities, such as merging the ACC and gyroscope [4], or the ACC, gyroscope and magnetometer [5], or the ACC and acoustic signals [6], to enhance the performance. However, not many works adopt the fusion of inertial and physiological sensors for the stable and accurate system.

Physiological sensors, such as photoplethysmography (PPG) and electrodermal activity (EDA), can capture physiological responses of an individual towards a stimuli event. While inertial sensors are useful in capturing the human activity with motions, some activities are either too similar in terms of physical movements or involving less motions. For example, driving and cycling are different activities, but both of them display similar physical movements. Those activities involve sitting state, but one is sitting on a bike and the other is sitting inside a car, which is difficult to be distinguished by inertial sensors. However, the physiological sensors can capture the inner physiological responses with respect to these two activities: driving requires high concentration that may vary the heart rate measured by the PPG, whereas cycling is an intense activity which may change the sweat condition captured by the EDA. There have been several attempts to develop the HAR system by fusing the inertial and physiological sensors. Early work in [7] integrates the ACC and PPG data, whereas [8] merges the PPG, ACC and gyroscope.

Most of these works rely on hand-crafted features to recognize human activities. G. Biagetti et al. [7] uses the singular value decomposition and truncated Karhunen-Loeve transform with the Bayesian classifier to distinguish the activities, whereas [8] employs a wavelet packet transform with the Random Forest (RF) to build the HAR system. While deep learning [9, 10, 11] and hierarchical classification [12] have been exploited in wearable-based HAR with inertial sensors, there are no works that combine both ideas with a fusion of physiological and inertial signals to infer activities. Also, even if all these works show the promising results in their applications, the various simulation scenarios are not considered

The work was funded by the Huawei Technologies Canada.

at the same time which is important for real applications. The original contributions of this work are summarized as follows:

- Investigating the effectiveness of the ACC, PPG and EDA for HAR system. Compared to the ACC, the PPG and EDA are not well-known but some activities can be recognized better with them.
- Experimenting three different scenarios (All-access, Cross-subject, Cross-activity) to follow the real world applications. Latter two scenarios follow the cross-validation which excludes some subjects and activities during training and only considers them during testing.
- Developing the Hierarchical Deep Learning (HDL) model to build the stable and accurate HAR system. This model is based on hierarchical tree structure with deep learning networks.

2. DATABASE

In this work, PPG-DaLiA database [13] has been used since it contains multimodal sensing modalities from both inertial and physiological signals collected when performing the closed to real life activities. This database was recorded with two different devices: RespiBAN Professional (chest-worn) and Empatica E4 (wrist-worn). For HAR, we focus on the ACC, PPG and EDA from wrist-worn device because of great accessibility and applicability. In PPG-DaLiA dataset, there are 15 subjects and sampling rate is fixed at 32 Hz for the ACC, 64 Hz for the PPG and 4 Hz for the EDA. In addition, 8 different activities are considered which is covered in Table 1 and there is a transient period between the activities for preparing the next activity. Only subject 6 has the hardware issue which results in missing three activities: lunch, walking, working. We select this database since it has activities with similar and small motions (ex. sitting/working/lunch) which are hard to be classified well but valuable to evaluate our HAR system.

3. METHODOLOGY

3.1. Sensing Modality

We experiment three different modalities to investigate the effectiveness of recognizing the human activities. First, we consider the **ACC** [14] which is the most well-known data to detect human movements because it can measure the acceleration forces in three axes according to the motions from activities. Next, we use the **PPG** [15] which optically captures the changes in blood volume from heart activity. There are some activities (ex. sitting/working/lunch) which do not include diverse motions and thus, it is hard to be distinguished by the ACC. In this case, the PPG can be useful since there can be different heart activities. Last, we investigate the **EDA** [16] which is the variation of the electrical conductance of the skin according to sweat condition. As the PPG, EDA is not considerably researched for HAR but it could be helpful to classify human activities with different sweat secretion.

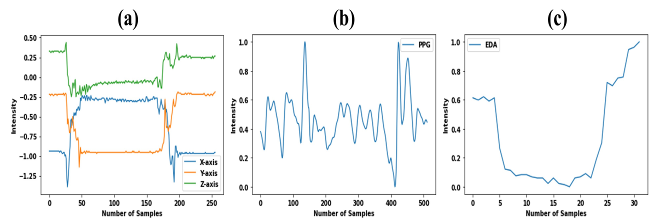

Fig. 1: Example of each modality with sliding window. (a-c) ACC, PPG, EDA. The length of window is different because of different sampling (ACC: 32 Hz, PPG: 64 Hz, EDA: 4 Hz).

Table 1: Details of all activities in PPG-DaLiA database.

Activity	Detail	Duration (min)
Sitting	Sitting still with reading.	10
Stairs	Climbing up and down.	5
Table Soccer	Playing 1 versus 1.	5
Cycling	2 km length with varying road conditions.	8
Driving	Performing on streets and country roads.	15
Lunch	Queuing and fetching food, eating and talking at the table.	30
Walking	Walking back from the canteen to the office.	10
Working	Mainly, working on computer.	20

The main reason to use three modalities is to build the robust and accurate HAR system even in activities with small motions. All these modalities are currently collected from wearable device and thus, they are suitable to apply in the real world. We employ the sliding window technique [17] over each modality with 8 seconds length and 2 seconds shift [18] which is empirically found to be proper for our application. The overlapped window is applied for increasing the size of data for training and testing, and for closer exploration of the temporal changes [19]. Figure 1 demonstrates the examples of data with sliding window where the PPG and EDA are normalized and resized to make all modalities in similar range to concentrate on the patterns and the shape of signals. Later, the generated inputs from window are utilized in the classification model to distinguish the activities.

3.2. HAR Models

Four models are considered to find the best suitable approach for HAR where key hyperparameters are covered in Table 2.

HDL Model: We develop this model which is composed of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) with hierarchical tree structure. The hierarchical structure employs the stack of models to provide the specific understanding at each level of data hierarchy which helps to improve accuracy [12]. Here, CNN gives the feature representations of the modalities and then, LSTM

Table 2: Key hyperparameters in all models.

	Hyperparameters	Values
RF, XGBoost	Number of trees	200
	Maximum depth	13
	Learning rate (Only for XGBoost)	0.3
HHAR-net	Dimensionality of the first and second layers	128, 64
	Dropout rate for the first layer	0.3
	Learning rate (ADAM optimization)	0.001
HDL	Width x Height x Depth of kernels in every CNNs	9x1x64
	Hidden unit of every LSTMs	256
	Dropout rate for every CNNs	0.5
	Learning rate (RMSprop optimization)	0.001

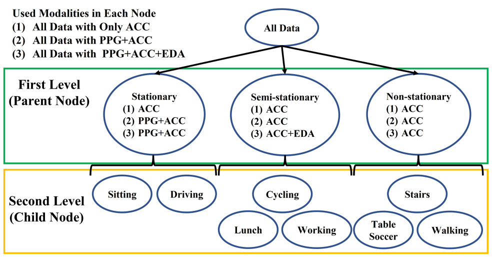

Fig. 2: The tree structures in HDL model. There are different used modalities in each node ((1)-(3)) depending on the contained modalities in the system. The reason for these differences are to boost the stability and accuracy of system.

models the temporal dynamics of that representation [20, 21, 22]. As Figure 2, there are two level classifications where the first level discriminates the stationary, semi-stationary and non-stationary activities, and the second level finds the specific activities in each parent node. Assigned activities in each node are explained in Figure 2. In parent level, each node has 4 CNNs with 3 LSTMs while in child level, each node consists of 3 CNNs with 2 LSTMs. Parent node has more layers since it has data with more diverse labels. Dropout is applied in every layers while ReLu is only used in CNNs.

Baseline Models: We compared with conventional models suitable for resource-constrained device that we consider.

First, we experiment the **RF** which builds the multiple decision trees and merges them to get an accurate and robust predictions. Our RF model follows the similar structure as [17]. Second, we consider the **eXtreme Gradient Boosting (XGBoost)** which is an elaborate version of the RF by adding a gradient boosting idea. This model is chosen to understand the ability of recent developed tree model for HAR. Last, we adopt the **HHAR-net** which applied the hierarchical classification with the NN [12]. This approach has a hierarchical tree structure as our HDL model and thus, we can see the effect of using the CNN and LSTM. In addition, [12] only considered the activities with obvious movements while HDL includes the semi-stationary categorization to improve the performances for activities like eating and working.

Table 3: Details about three different scenarios. In all cases, there is no overlap between train and test sets.

	Details
All-access	Extract train (80%) and test (20%) sets from each activity in all subjects. Access all activities and subjects for training.
Cross-subject	12 subjects for training and 3 subjects for testing. Sequentially process (i.e. first test set is subject 1,2,3 and the second one is 4,5,6 while remained subjects are for train set). Average 5 trials (cross-validation).
Cross-activity	Randomly exclude two activities in each subject for testing while others for training. Average 4 trials (cross-validation).

Table 4: Overall performances according to different modalities, models and scenarios. A, P and E mean the ACC, PPG and EDA. The performances are described as ACR (F1 score).

	All-access			**Cross-subject**			**Cross-activity**		
	A	**P+A**	**P+A+E**	**A**	**P+A**	**P+A+E**	**A**	**P+A**	**P+A+E**
RF	94.13% (78.71%)	94.16% (78.25%)	93.95% (77.84%)	88.91% (56.99%)	88.85% (56.7%)	89.13% (58%)	87.23% (50.65%)	87.1% (49.98%)	87.27% (50.83%)
XGBoost	95.82% (84.6%)	95.72% (84.35%)	95.63% (84.02%)	89.62% (59.86%)	90.07% (61.9%)	90.64% (65.09%)	87.92% (53.41%)	88% (54.36%)	88.6% (57.49%)
HHAR-net	92.56% (73.1%)	92.42% (72.39%)	92.45% (71.85%)	88.05% (53.76%)	88.46% (56.11%)	89.27% (58.95%)	86.75% (48.67%)	86.86% (49%)	87.07% (49.80%)
HDL	95.92% (85.95%)	**97.16% (90.63%)**	95.74% (86.47%)	91.05% (68.76%)	**92.23% (73.48%)**	92.01% (71.91%)	89.79% (61.99%)	**90.18% (64.07%)**	89.84% (62.62%)

Table 5: Comparison of performances in same database. F1 is not shown because [17] did not cover the overall F1 score.

	[17]	**Ours**	
Scenario	All-access	**All-access**	
Modality	electrocardiogram, ACC, PPG, EDA, respiration, body temperature	**PPG, ACC**	
Input	8 seconds window with Fourier transform	**8 seconds window**	
Model	RF	**RF**	**HDL**
ACR	92.8%	94.16%	**97.16%**

4. EXPERIMENTS AND RESULTS

In this work, we cover the experimental results according to different modalities, models and scenarios. All the simulations are considered as multi-class classification since we are concurrently discriminating all activities.

4.1. Experimental Scenarios, Modalities and Metrics

Three different scenarios (Table 3) are considered, namely All-access, Cross-subject and Cross-activity. In All-access case, we access all subjects and activities to train the model while in other two cases, we exclude some subjects and activities during training and only use them for testing. Latter two scenarios follow the realistic scenarios because our HAR system can be evaluated in the unseen subjects and activities. Nevertheless, All-access is useful to understand the effect of overlapped window when comparing with the other scenarios. Previously, Cross-activity is less considered even if it has its practical value. For example, in PPG-DaLiA [13], subject 6

Table 6: Comparison of results in different databases. [12, 23] used a dataset [24] from the smartphone and smartwatch.

	[23]	[12]	**Ours**		
Activities	standing/sitting/sleeping/ exercise/walking	standing/sitting/lying down/ running/walking/bicycling	**8 activities in Table 1**		
Scenario	All-access	All-access	**All-access**		
Modality	ACC, gyroscope, magnetometer, watch compass	ACC, gyroscope, watch compass, location, audio, phone state	**PPG, ACC**		
Input	Features from 20 seconds window	Features from 20 seconds window	**8 seconds window**		
Model	XGBoost	HHAR-net	**XGBoost**	**HHAR-net**	**HDL**
ACR (F1)	86.8% (60%)	92.8% (87.42%)	95.72% (84.35%)	92.42% (72.39%)	**97.16% (90.63%)**

Table 7: Results in each activity according to models and scenarios with the PPG+ACC. Each cell denotes as ACR (F1 score).

	Cross-subject				Cross-activity			
	RF	XGBoost	HHAR-net	HDL	RF	XGBoost	HHAR-net	HDL
Sitting	88.88% (23.47%)	91.04% (40.18%)	90.87% (48.2%)	**92.93% (62.51%)**	88.22% (23.17%)	89.96% (33.34%)	90.12% (40.02%)	**92.59% (55.63%)**
Stairs	94.19% (57.95%)	94.71% (62.78%)	94.15% (54.5%)	**97.54% (81.82%)**	92.31% (47.38%)	92.85% (52.15%)	91.95% (44.18%)	**95.14% (68.15%)**
Table Soccer	97.18% (68.19%)	97.61% (70.23%)	96.59% (62.33%)	**98.13% (78.7%)**	97.08% (65.84%)	97.42% (69.04%)	96.24% (57.62%)	**97.49% (70.86%)**
Cycling	97.48% (83.22%)	98.41% (86.2%)	97.13% (78.72%)	**99.26% (93.97%)**	97.53% (83.48%)	98.3% (88.47%)	96.85% (77.69%)	**99.23% (94.91%)**
Driving	88.06% (59.8%)	89.35% (64.09%)	86.16% (52.22%)	**93.71% (77.52%)**	85.71% (53.53%)	86.79% (57.79%)	85.07% (50.79%)	**92% (73.07%)**
Lunch	68.92% (54.17%)	72.45% (58.76%)	68.53% (53.02%)	**77.74% (62.84%)**	65.42% (48.62%)	67.53% (51.35%)	65.87% (47.75%)	**72.97% (56.17%)**
Walking	92.99% (59.07%)	93.49% (63.19%)	92.9% (59.8%)	**96% (80.17%)**	89.79% (47.99%)	90.64% (53.33%)	89.97% (49.21%)	**92.44% (59.17%)**
Working	83.08% (47.75%)	83.49% (49.75%)	81.31% (40.07%)	**82.49% (50.34%)**	80.77% (29.87%)	80.49% (29.38%)	78.84% (24.71%)	**79.62% (34.62%)**

Table 8: Confusion matrix according to different modalities with the HDL in Cross-subject scenario.

ACC_HDL

Predicted \ Actual	Sitting	Stairs	Soccer	Cycling	Driving	Lunch	Walking	Working
Sitting	133	1	0	1	18	94	1	70
Stairs	2	177	0	0	0	8	36	1
Soccer	0	2	106	1	6	13	2	1
Cycling	0	1	1	214	1	2	0	1
Driving	6	1	4	7	353	41	2	20
Lunch	63	6	36	5	46	546	30	207
Walking	0	24	1	0	0	18	237	3
Working	99	1	3	2	29	178	2	262

PPG+ACC_HDL

Predicted \ Actual	Sitting	Stairs	Soccer	Cycling	Driving	Lunch	Walking	Working
Sitting	195	0	1	1	9	49	2	31
Stairs	0	172	0	0	0	7	27	1
Soccer	0	1	118	1	4	17	2	1
Cycling	0	1	1	214	1	4	0	1
Driving	19	1	4	7	365	38	4	35
Lunch	46	4	24	6	50	578	25	209
Walking	2	32	2	0	0	21	249	3
Working	40	2	2	1	23	187	2	284

PPG+ACC+EDA_HDL

Predicted \ Actual	Sitting	Stairs	Soccer	Cycling	Driving	Lunch	Walking	Working
Sitting	184	0	1	0	7	53	2	25
Stairs	1	172	0	0	0	6	41	1
Soccer	1	1	110	2	6	11	2	1
Cycling	0	1	1	211	0	4	0	4
Driving	23	0	5	6	341	23	1	20
Lunch	50	14	31	8	65	594	39	197
Walking	1	24	0	0	0	15	217	3
Working	42	2	3	3	33	194	8	314

does not have three activities because of hardware issue, and the Cross-activity scenario can simulate this case.

In addition, we tested unimodal and multimodal HAR systems. For unimodal, only ACC is utilized while for multimodal, PPG+ACC or PPG+ACC+EDA is used. In both systems, inputs are concatenated in 1 channel (ex. ACC: [256, 3]⇒[768, 1]) for the baseline models while they are stacked on multi-channels (ex. PPG+ACC: [256, 4]) for the HDL.

Three different metrics are applied to evaluate the performance, namely accuracy (ACR), F1 score and confusion matrix. ACR is the number of correctly classified data according to the total number of data which is suitable in even data distribution. In the real world, this is unusual and we additionally consider F1 score, the weighted average of precision and recall. This metric is more appropriate in uneven data distribution. Confusion matrix represents the predicted class of data with respect to the actual class. It is useful to analyze the distribution of miss-classified data in each activity.

4.2. Results and Discussion

Overall Evaluation: Table 4 shows the overall performances with different modalities, models and scenarios. Compare to the All-access, other two scenarios show the decreased performances which reveals the difficulty of recognizing the unseen subjects and activities. This aspect also confirms that the overlapped window has minimal effects on training the model. We can notice that our developed HDL enhances the results, especially when combined with the PPG+ACC. The fusion of the EDA shows a substantial improvement for cross-validation scenarios in baseline models. In Table 5, we compare the performances with different approach in same database. Unlike [17], we only consider the PPG and ACC but get a better result with same RF model. Thus, including more modalities cannot promise the increase in performance. In Table 6, we compare the results with same models in different databases. Although it is not perfect way to compare the performances but we can understand whether our approach is compatible to the recent works. Even with less modalities, XGBoost with our approach shows better result to distinguish more diverse activities. HHAR-net with our method gives worse performance than [12], but we consider more various activities and use a short length of window, which requires the short time to recognize the activity. In both Table 5 and 6, it is obvious that our HDL gives the best performances. To understand the model's ability deeply, we introduce Table 7 which represents the result in each activity with the PPG+ACC in Cross-subject and Cross-activity cases. From Table 7, HDL gives the best performance in terms of F1 score in all activities even with different scenarios. Thus, our proposed HAR system is stable and accurate in all activities.

Cross-subject Evaluation: In Table 8, we cover the confusion matrix with different modalities to analyze the effect of each modality. When including the PPG with the ACC, most activities are classified better, except for stair and cycling conditions. This result comes from the ability of PPG to capture different heart activities in each condition. When we compare the middle and right matrices, the EDA gives better distinguishable features for lunch and working conditions, meaning it is useful to address the semi-stationary activities.

Cross-activity Evaluation: In a practical scenario, some subjects might fail to provide the required activities for training purposes. This is especially true with our chosen database where subject 6 failed to provide the data for 3 activities. Table 7 verifies that most of the trained models attain reasonable performances when encountered such scenario and our HDL achieves the best result.

5. CONCLUSION

This work focuses on the fusion of inertial and physiological sensors from smartwatches to build a HAR system. To improve the robustness, we suggest a HDL model evaluated on three realistic scenarios. Our developed system achieves the eminent results which outperform other state-of-art works.

6. REFERENCES

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys Tutorials*, pp. 1192–1209, 2013.

[2] P. Bharti, A. Panwar, G. Gopalakrishna, and S. Chellappan, "Watch-dog: Detecting self-harming activities from wrist worn accelerometers," *IEEE Journal of Biomedical and Health Informatics*, pp. 686–696, 2018.

[3] M. Abbas and R. L. B. Jeannès, "Characterizing peaks in acceleration signals–application to physical activity detection using wearable sensors," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 12384–12395, Oct 2020.

[4] M. Webber and R. F. Rojas, "Human activity recognition with accelerometer and gyroscope: A data fusion approach," *IEEE Sensors*, pp. 16979–16989, Aug 2021.

[5] A. Nazábal and A. Artés-Rodríguez, "Discriminative spectral learning of hidden markov models for human activity recognition," in *2015 IEEE ICASSP*. IEEE, 2015, pp. 1966–1970.

[6] J.A. Ward, P. Lukowicz, G. Troster, and T.E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, Oct 2006.

[7] G. Biagetti, P. Crippa, L. Falaschetti, S. Orcioni, and C. Turchetti, "Human activity recognition using accelerometer and photoplethysmographic signals," in *IDT*. Springer, 2017, pp. 53–62.

[8] N. S. Pathan, M. T. F. Talukdar, M. Quamruzzaman, and S. A. Fattah, "A machine learning based human activity recognition during physical exercise using wavelet packet transform of ppg and inertial sensors data," in *EICT*. IEEE, 2019, pp. 1–5.

[9] S. A. Rokni, M. Nourollahi, and H. Ghasemzadeh, "Personalized human activity recognition using convolutional neural networks," in *AAAI*, 2018, vol. 32.

[10] H. Ma, W. Li, X. Zhang, S. Gao, and S. Lu, "Attnsense: Multi-level attention mechanism for multimodal human activity recognition.," in *IJCAI*, 2019, pp. 3109–3115.

[11] A. Akbari and R. Jafari, "An autoencoder-based approach for recognizing null class in activities of daily living in-the-wild via wearable motion sensors," in *ICASSP*. IEEE, 2019, pp. 3392–3396.

[12] M. Fazli, K. Kowsari, E. Gharavi, L. Barnes, and A. Doryab, "Hhar-net: Hierarchical human activity recognition using neural networks," in *IHCI*. Springer, 2020, pp. 48–58.

[13] A. Reiss, I. Indlekofer, P. Schmidt, and K. V. Laerhoven, "Deep ppg: Large-scale heart rate estimation with convolutional neural networks," *Sensors*, vol. 19, no. 14, pp. 3079, 2019.

[14] A. Jordao, L. A. B. Torres, and W. R. Schwartz, "Novel approaches to human activity recognition based on accelerometer data," *Signal, Image and Video Processing*, vol. 12, no. 7, pp. 1387–1394, 2018.

[15] E. Brophy, J. J. D. Veiga, Z. Wang, A. F. Smeaton, and T. E. Ward, "An interpretable machine vision approach to human activity recognition using photoplethysmograph sensor data," *arXiv:1812.00668*, 2018.

[16] Y. Zhang, M. Haghdan, and K. S. Xu, "Unsupervised motion artifact detection in wrist-measured electrodermal activity data," in *ISWC*, 2017, pp. 54–57.

[17] A. P. Psathas, A. Papaleonidas, and L. Iliadis, "Machine learning modeling of human activity using ppg signals," in *ICCCI*. Springer, 2020, pp. 543–557.

[18] Z. Zhang, Z. Pi, and B. Liu, "Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Transactions on biomedical engineering*, vol. 62, no. 2, pp. 522–531, 2014.

[19] T. Mahmud, S. S. Akash, S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Human activity recognition from multimodal wearable sensor data using deep multi-stage lstm architecture based on temporal feature aggregation," in *MWSCAS*. IEEE, 2020, pp. 249–252.

[20] M. Boukhechba, L. Cai, C. Wu, and L. E. Barnes, "Actippg: using deep neural networks for activity recognition from wrist-worn photoplethysmography (ppg) sensors," *Smart Health*, vol. 14, pp. 100082, 2019.

[21] R. Mutegeki and D. S. Han, "A cnn-lstm approach to human activity recognition," in *ICAIIC*. IEEE, 2020, pp. 362–366.

[22] S. Mekruksavanich and A. Jitpattanakul, "Smartwatch-based human activity recognition using hybrid lstm network," in *2020 IEEE Sensors*. IEEE, 2020, pp. 1–4.

[23] P. Tarafdar and I. Bose, "Recognition of human activities for wellness management using a smartphone and a smartwatch: a boosting approach," *Decision Support Systems*, vol. 140, pp. 113426, 2021.

[24] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE pervasive computing*, vol. 16, no. 4, pp. 62–74, 2017.