

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/327344984>

A Machine Vision Approach to Human Activity Recognition using Photoplethysmograph Sensor Data

Article · August 2018

CITATIONS

0

READS

73

4 authors:



Eoin Brophy

Dublin City University

3 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)



Zhengwei Wang

Dublin City University

9 PUBLICATIONS 5 CITATIONS

[SEE PROFILE](#)



Jose Juan Dominguez Veiga

Dublin City University

8 PUBLICATIONS 36 CITATIONS

[SEE PROFILE](#)



Tomas E Ward

Dublin City University

253 PUBLICATIONS 2,741 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Development & Evaluation of Wearable Sensor Based Exercise Biofeedback Systems [View project](#)



Rapid serial visual presentation based brain computer interfaces [View project](#)

A Machine Vision Approach to Human Activity Recognition using Photoplethysmograph Sensor Data

Eoin Brophy
Department of Electronic Engineering
Maynooth University
Maynooth, Kildare
eoin.brophy.2013@mumail.ie

Zhengwei Wang
Insight Centre for Data Analytics
Dublin City University
Glasnevin, Dublin
zhengwei.wang22@mail.dcu.ie

José Juan Dominguez Veiga
Insight Centre for Data Analytics
Dublin City University
Glasnevin, Dublin
jose.dominguezveiga3@mail.dcu.ie

Tomas E. Ward
Insight Centre for Data Analytics
Dublin City University
Glasnevin, Dublin
tomas.ward@dcu.ie

Abstract— *Human activity recognition (HAR) is an active area of research concerned with the classification of human motion. Cameras are the gold standard used in this area, but they are proven to have scalability and privacy issues. HAR studies have also been conducted with wearable devices consisting of inertial sensors. Perhaps the most common wearable, smart watches, comprising of inertial and optical sensors, allow for scalable, non-obtrusive studies. We are seeking to simplify this wearable approach further by determining if wrist-mounted optical sensing, usually used for heart rate determination, can also provide useful data for relevant activity recognition. If successful, this could eliminate the need for the inertial sensor, and so simplify the technological requirements in wearable HAR. We adopt a machine vision approach for activity recognition based on plots of the optical signals so as to produce classifications that are easily explainable and interpretable by non-technical users. Specifically, time-series images of photoplethysmography signals are used to retrain the penultimate layer of a pretrained convolutional neural network leveraging the concept of transfer learning. Our results demonstrate an average accuracy of 75.8%. This illustrates the feasibility of implementing an optical sensor-only solution for a coarse activity and heart rate monitoring system. Implementing an optical sensor only in the design of these wearables leads to a trade off in classification performance, but in turn, grants the potential to simplify the overall design of activity monitoring and classification systems in the future.*

Keywords— *deep learning, activity recognition, biomedical, photoplethysmography*

I. INTRODUCTION

Due to the ubiquitous nature of inertial and physiological sensors in phones and fitness trackers, human activity recognition (HAR) studies have become more frequent [1],[2]. Benefits of HAR include rehabilitation for recovering patients [3], monitoring of the elderly and vulnerable or advancements in human-centric applications [4].

Photoplethysmography (PPG) is an optical technique used to measure volume changes of blood in the microvascular tissue. PPG is capable of measuring heart rate by detecting the

amount of light reflected/absorbed in red blood cells as this varies with the cardiac cycle. The light that is reflected is read by an ambient light sensor which then has its output conditioned, so a pulse rate can be determined from the module. The pulse rate is obtained from analysis of the small ac component (which arises from the pulsatile nature of blood flow) superimposed on a large dc component caused by the constant absorption of light [5].

For usability reasons the wrist is a common site for wearables used in health and fitness contexts [6]. Most smartwatches are equipped with an optical PPG sensing device capable of calculating pulse rate. Difficulties arising in obtaining a robust physiologic output signal from a PPG can be caused by motion artefact due to the changes in optical path length associated with disturbance of the source-detector configuration. This disturbance is introduced by haemodynamic effects and gross motor movements of the limbs [7]. This can often lead to an incorrect reading of the pulse rate signal. Reduction in motion artefacts can be achieved using a range of techniques, from aggressive filtering to adaptive methods based on a measure of the artefact source from an accelerometer-based measurement [8].

In this study we sought to exploit the motion artefact and infer human activity from the PPG signals collected at the wrist. Our hypothesis was that there is sufficient information in the disturbance induced in the source-detector path to distinguish different activities through the use of a machine learning approach. In recent years, capabilities of machine learning methods in the field of image recognition has increased dramatically [9]. Utilising these advancements in image recognition will allow for simplification of wearables involved in HAR. We chose an image-based approach to the machine learning challenge as this work is part of a larger scoped effort to develop easily deployed artificial intelligence which can be used and interpreted by end users who do not have deep levels of signal processing expertise – [10].

A. Related Work

Convolutional neural networks (CNNs) have been used since the 1990's and were designed to mimic how the visual cortex of the human brain processes and recognises images [11]. CNNs extract salient features from images at various layers of the network. They allow implementation of high-accuracy image classifiers given the correct training without the need for in-depth feature extraction knowledge.

Current state of the art activity recognition systems are camera-based. These can detect motion with little processing but require a lot of processing to recognise specific activities. Cameras also introduce scalability issues along with being intrusive by nature [12]. Inertial sensing is another popular method used in HAR. To achieve the high accuracies of the inertial sensing systems shown in [6], a system consisting of multiple sensors is required, again compromising of functionality and scalability issues. The associated signal processing is not trivial and singular value decomposition (SVD), truncated Karhunen-Loève transform (KLT) [13], Random Forest (RF) and Support Vector Machines (SVM) are examples of feature extraction and machine learning methods that can be applied to human activity recognition.

Inertial sensor data paired with PPG are amongst the most suitable sensors for activity monitoring as they offer effective tracking of movement actions as well as relevant physiological parameters such as heart rate. They also have the benefit of being easy to deploy. Mehrang *et al.* used RF and SVM machine learning methods for a HAR classifier on combined accelerometer and PPG data achieving an average recognition accuracy of 89.2% and 85.6% respectively [14].

The average classification accuracy of the leading modern feature extraction and machine learning methods for singular or multiple accelerometers sensors range from 80% to 99% [6]. However, this can require up to 5 accelerometers located at various positions of the body.

II. METHODOLOGY

A. Data Collection

The design of the data collection experiment was conducted by Deleram Jarchi and Alexander J. Casson and is freely available from Physionet [15], [16]. PPG recordings were taken from 8 patients (5 female, 3 male) aged between 22-32 (mean age of 26.5), during controlled exercises on a treadmill and an exercise bike. Data was recorded using a wrist-worn PPG sensor attached to the Shimmer 3 GSR+ unit [17] for an average period of 4-6 minutes with a maximum duration of 10 minutes. A frequency of 256 Hz was used to sample the physiological signal. Participants used a treadmill to run and walk and an exercise bike on high/low resistance to simulate 4 variations of exercises. The data analysed was the raw signal with minimal filtering to keep the motion artefact from the wrist-worn sensor present. Each individual was allowed to set the intensity of their exercises as they saw fit and every

exercise began from rest.

The four exercises were broken down as; walk on a treadmill, run on a treadmill, low resistance exercise bike and high resistance exercise bike. For the walk and run exercises the raw PPG signals required no filtering other than what the Shimmer unit provides. The cycling recordings were low-pass filtered using Matlab with a 15 Hz cut-off frequency to remove the high-frequency noise.

B. Data Preparation

The PPG data signal was downloaded using the PhysioBank ATM and plotted in Matlab. The signals were segmented into smaller time-series windows of 8-second intervals, these intervals were chosen to match the time windows used in [18] which acts as a benchmark for this study. A rectangular windowing function was created to step through the data every 2 seconds and save a new plot of 8 seconds of data. It is worth re-emphasising that a machine vision approach is being taken here – the input data to the classifier is not time series vectors but actual images. These images correspond to minimalist plots of the 8-second window produced in Matlab. Therefore, the script prints and saves figures and removes all axis labels, legends and grid ticks (removing non-salient features), saving each figure as a 1201x901 JPEG file. A total of 3321 images were created, of which 80% (2657) were used for retraining, 10% (332) for validation and 10% (332) for testing.

These .jpg files were stored in a directory hierarchy based on the movement carried out. Four subdirectories of possible classifiers were created; run, walk, high resistance bike and low resistance bike. Contained within each subdirectory was the images of the PPG signal plotted in Matlab. In Fig. 1 below an example of each activity can be seen.

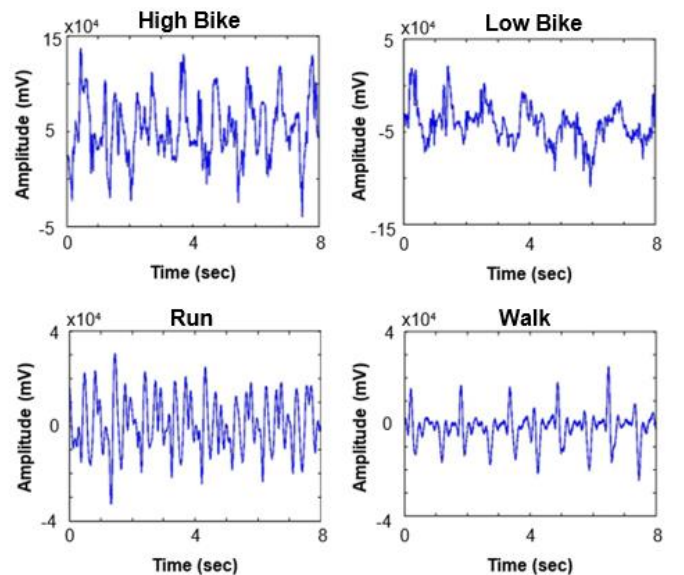


Fig. 1 Sample of PPG images for each activity

C. The Network Infrastructure

Upon completion of the data preparation the convolutional neural network could then be retrained. Building a neural network from the ground up is far from a trivial task, it requires multilayer implementation for even a simple perceptron [11] which needs optimization of tens of thousands of parameters for even a trivial task such as handwritten digit classification [19]. Instead of spending a lot of time building a neural network from scratch, the authors used Inception which can be implemented with the TensorFlow framework [20].

TensorFlow, a deep learning framework was used by the authors for transfer learning. A Python distribution, Anaconda, was the container in which TensorFlow was installed. Transfer learning is the idea of using a pretrained convolutional neural network and retraining the penultimate layer that does the classification before the output. This type of learning is ideal for this study due to the small dataset [21]. The results of the retraining process can be viewed using the suite of visualisation tools on TensorBoard [22].

Recognition of specific objects from millions of images requires a model with a large learning capacity. CNNs are particularly suitable for image classification because convolution leverages three important properties that can help improve a machine learning system: sparse interaction, parameter sharing and equivariant representations [23]. These properties enable CNN to detect small, meaningful features from the input image and reduce the storage requirements of the model comparing to the traditional densely connected neural networks.

CNNs are tuneable in their depth and breadth which means they can have much simpler architecture and in turn making them easier to train in comparison to other feedforward neural networks [9]. The CNN recognises patterns across space and in this case, recognises objects from images. However, if overfitted, CNNs remember recurring patterns in the background of each image it was trained on and using that to match labels with objects. It produces good results but when new images are introduced to the network it fails due to an inability to capture general characteristics of the training dataset.

D. Retraining and Using the Network

To retrain the network, the same approach carried out by Dominguez Veiga *et. al* was taken [10]. A pretrained CNN, Google's Inception-v3, was used which has been trained on ImageNet, a database of over 14 million images and will soon be trained on up to 50 million images [24]. The retraining CNN tutorial from Google [25] was followed and allowed the authors to retrain just the last layer of the network using their own set of images.

The retraining process can be fine-tuned through hyperparameters which allows for optimisation of the training outcome. During the training for this paper the default parameters were used except for the training steps which were changed from the default of 4,000 steps to 10,000 steps.

Selection of this number of iterations allowed for the loss function (cross-entropy) to be sufficiently converged, avoiding overfitting. Equation (1) shows the formula for cross-entropy where M is the number of classes, y is a binary indicator (0 or 1) if label c is the correct classification for prediction o and p is the predicted probability o is of class c [26].

$$\sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (1)$$

III. RESULTS

The results for each of the 10,000 steps were output into the two files 'output_graph.pb' and 'output_labels.txt', this allows the results to be viewed graphically on TensorBoard. Output in TensorBoard was the training and validation accuracy (Fig. 2) along with the cross-entropy (Fig. 3). The graphs have been curve fitted using a smoothing function.

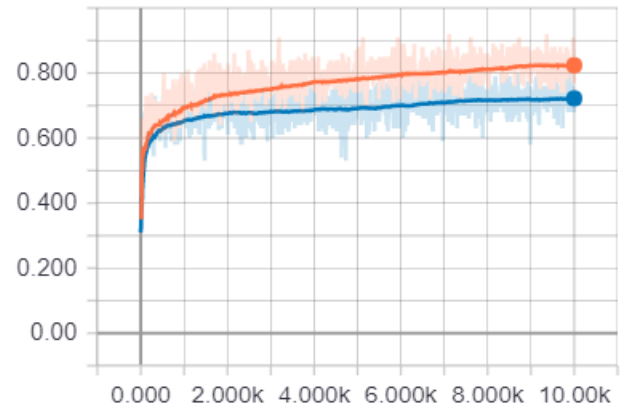


Fig. 2 Training (Orange) vs. Validation (Blue) Plot

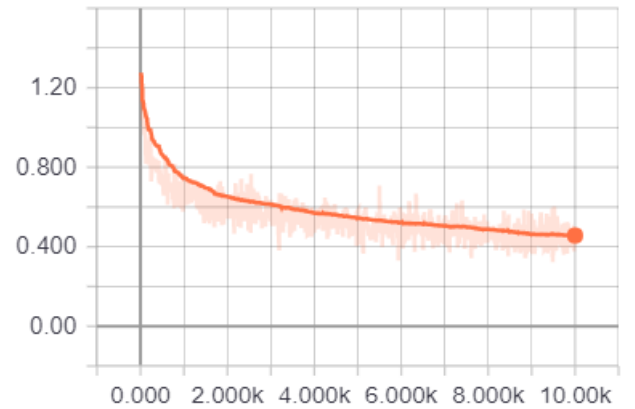


Fig. 3 Cross-Entropy Plot

The loss function is continuously reducing on every iteration of the training step which shows that the model prediction is getting closer to the true distribution with each step. The final accuracy was shown to be 75.8% and as can be seen in Fig. 4; the confusion matrix, demonstrates the validation accuracy for correctly classifying the test set of images.

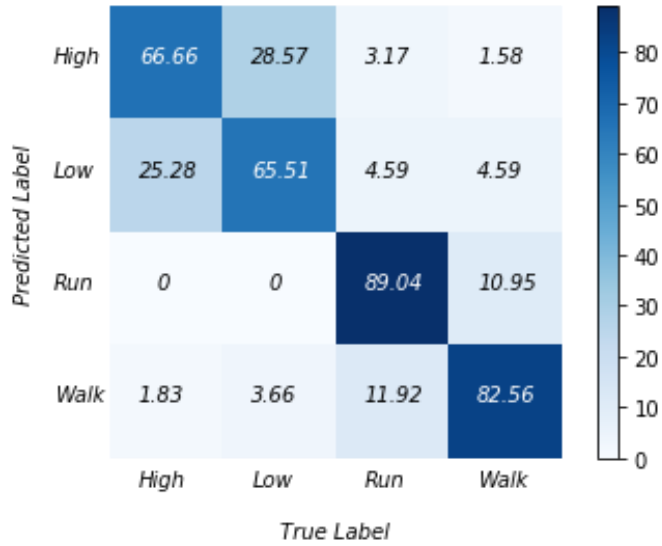


Fig. 4 Confusion matrix for transfer learning approach

The confusion matrix for the deep learning approach graphically demonstrates some of the issues classifying the high-resistance bike exercise, where it was misclassified as low-resistance 28.57% of the time. The other activities had a maximum HAR misclassification percentage of 25.28% and a minimum of 0%.

Fig. 6 shows two examples of each class that were misclassified as other activities. Participants were encouraged to go at their own pace. One participant's walk may be as fast as another's run which could be an issue for the classifier. However, based on the plots shown in Fig. 6, the errors may have been from a loose wrist strap or excessive movement of the arms. In some circumstances, the signal can be seen to be cut off which indicates gross movement of the limbs for those time instances [27].

A study conducted by Giorgio Biagetti *et al.* used feature extraction and reduction with Bayesian classification on time-series signals [18]. Their technique focused on using singular value decomposition and truncated Karhunen-Loève transform. Their study used the same time-series dataset and was designed to present an efficient technique for HAR using PPG and accelerometer data. Below, in Fig. 5 the confusion matrix for their results can be seen, using *just* the PPG signal for feature extraction and classification.

The feature extraction approach for determining HAR using just the PPG yields an overall accuracy of 44.7%. This shows a reduced classification performance versus the deep learning

method employed in this paper. Although our transfer learning achieved greater accuracy of over 30 percentage points (75.8% vs. 44.7%), Biagetti *et al.* in the same paper then combined the PPG and accelerometer data to bring their classifier accuracy rate to 78%. We are able to produce very acceptable accuracy without the use of an accelerometer, i.e. through the optical signal only.

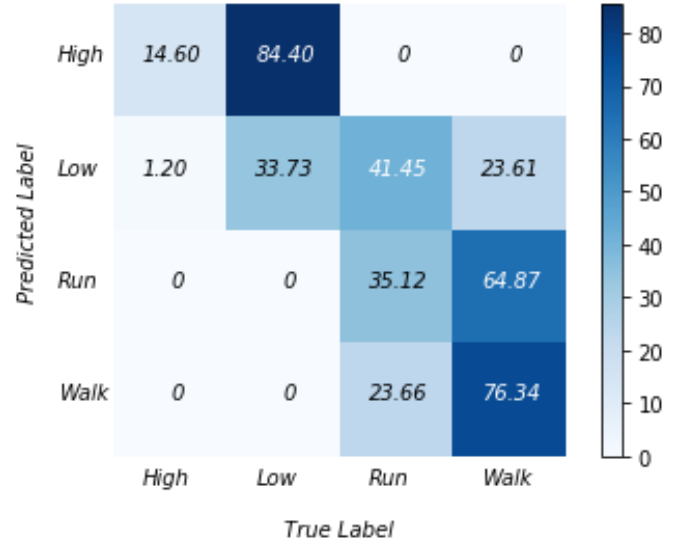


Fig. 5 Confusion matrix of feature extraction approach

IV. DISCUSSION

A. Principal Findings

Applying transfer learning to the PPG dataset leads to a classification accuracy of 75.8% (2607/3321). This is very close to the combined PPG and accelerometer data for HAR using SVD and KVL (75.8% vs. 78%) and much better than the PPG only result (75.8% vs. 44.7%) [18]. This is a competitive result and suggests that simpler wearables based on optical measurements only could yield much of the functionality achievable with more sophisticated, existing multi-modal devices. Of course, the addition of an inertial sensor will always produce more information and therefore more nuanced activity recognition. However, for the types of activity recognition commonly sought in clinical, health and fitness applications a surprisingly good performance can be extracted from a very simple optical measurement.

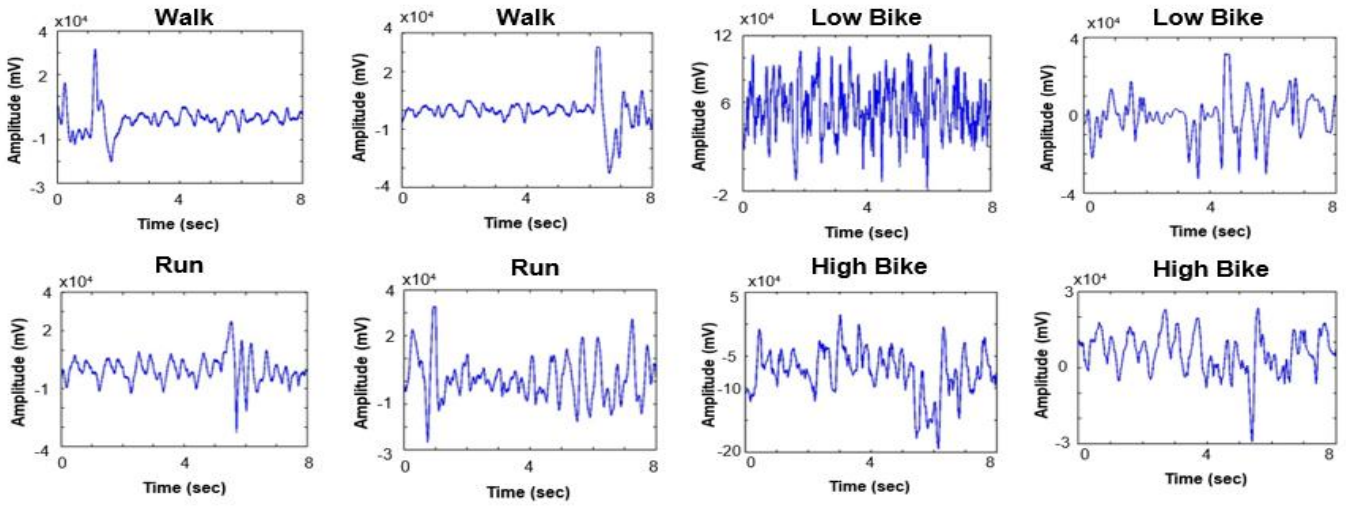


Fig. 6 Sample of eight misclassified images

B. Limitations

Better understanding of the hyperparameters may lead to higher average classification accuracy than the one achieved in this paper (75.8%). Given more knowledge of building complete neural networks from scratch against just retraining the last layer may also yield better results. The results generated in this paper are based on a classifier that was trained on data gathered from activities undertaken in an experimental setting. However, it has not been tested on everyday activities outside the laboratory environment.

V. CONCLUSION

Transfer learning for human activity recognition is a novel approach to extracting new information from wrist-worn PPG sensors which are conventionally used for heart rate monitoring. Signal processing studies using PPG sensors have found them to make but minimal contributions to improvement in classification accuracy for HAR [28]. This has led to a lack of independent studies using a standalone PPG sensor, most of which have been combined with an accelerometer. The capability of CNNs to create classifiers from detailed images allow our retrained model to be implemented successfully for HAR on images of PPG data [9].

Accuracy and simplicity of the retrained CNN has proven to be a great benefit of the deep learning approach adopted here. Users of this system do not need to possess a strong signal processing background to understand the approach and this allows the possibility that non-experts can develop their own HAR classification applications more readily.

Pathways for HAR using deep learning are beginning to be explored on a larger scale thanks to the simplicity of a transfer learning approach, cutting the development of a CNN from two weeks down to a few hours. This new method of feature classification will allow for the easier testing of hypotheses relating to HAR with wearables.

The presented process allows activity classification models to be constructed using PPG sensors only, potentially

eliminating the need for an inertial sensor set and simplifying the overall design of wearable devices.

VI. ACKNOWLEDGEMENT

This project was partly funded by the John Hume Studentship from Maynooth University. The Insight Centre for Data Analytics is supported by Science Foundation Ireland under grant number SFI/12/RC/2289.

VII. ABBREVIATIONS

CNN	convolutional neural network
GSR	galvanic skin response
HAR	human activity recognition
KLT	Karhunen-Loève transform
PPG	photoplethysmography
RF	random forest
SVD	singular value decomposition
SVM	support vector machines

REFERENCES

- [1] A. Filippoupolitis, B. Takand, and G. Loukas, "Activity Recognition in a Home Setting Using off the Shelf Smart Watch Technology," *Proc. - 2016 15th Int. Conf. Ubiquitous Comput. Commun. 2016 8th Int. Symp. Cybersp. Secur. IUCC-CSS 2016*, pp. 39–44, 2017.
- [2] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
- [3] E. S. Sazonov, G. Fulk, N. Sazonova, and S. Schuckers, "Automatic recognition of postures and activities in stroke patients," *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009*, pp. 2200–2203, 2009.
- [4] E. Kim, S. Helal, and D. Cook, "Human Activity Recognition and Pattern Discovery," *Pervasive Comput. IEEE*, vol. 9, no. 1, pp. 48–53, 2010.
- [5] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "A Novel Approach for Motion Artifact Reduction in PPG Signals Based on AS-LMS Adaptive Filter," *Instrum. Meas. IEEE Trans.*, vol. 61, no. 5, pp. 1445–57, 2012.
- [6] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010.
- [7] A. M. Tautan, A. Young, E. Wentink, and F. Wieringa, "Characterization and reduction of motion artifacts in photoplethysmographic signals from a wrist-worn device," *Proc.*

- Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015–Novem, pp. 6146–6149, 2015.
- [8] J. Allen, “Photoplethysmography and its application in clinical physiological measurement,” *Physiol. Meas.*, vol. 28, no. 3, 2007.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.
- [10] J. J. Dominguez Veiga, M. O’Reilly, D. Whelan, B. Caulfield, and T. E. Ward, “Feature-Free Activity Classification of Inertial Sensor Data With Machine Vision Techniques: Method, Development, and Evaluation,” *JMIR mHealth uHealth*, vol. 5, no. 8, p. e115, 2017.
- [11] S. Raschka and V. Mirjalili, *Python Machine Learning*. Packt Publishing, 2017.
- [12] A. Schmidt and K. Van Laerhoven, “How to build smart appliances?,” *IEEE Pers. Commun.*, vol. 8, no. 4, pp. 66–71, 2001.
- [13] M. A. O. Vasilescu, “Human motion signatures: analysis, synthesis, recognition,” *Int. Conf. Pattern Recognition, 2002. Proc.*, pp. 456–460 vol.3, 2002.
- [14] S. Mehrang *et al.*, “Human Activity Recognition Using A Single Optical Heart Rate Monitoring Wristband Equipped with Triaxial Accelerometer,” in *European Medical and Biological Engineering Conference*, 2017, pp. 587–590.
- [15] D. Jarchi and A. Casson, “Description of a Database Containing Wrist PPG Signals Recorded during Physical Exercise with Both Accelerometer and Gyroscope Measures of Motion,” *Data*, vol. 2, no. 1, p. 1, 2016.
- [16] A. L. Goldberger *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [17] Shimmer, “SHIMMER3 GSR + Unit,” 2015. [Online]. Available: <http://www.shimmersensing.com/shop/shimmer3-wireless-gsr-sensor#video-tab>. [Accessed: 01-Feb-2018].
- [18] G. Biagetti, P. C. B. L. Falaschetti, S. Orcioni, C. Turchetti, and D. I. I. Dipartimento, “Human Activity Recognition Using Accelerometer and Photoplethysmographic Signals,” in *Intelligent Decision Technologies*, 2018, vol. 73.
- [19] TensorFlow, “A Guide to TF layers Building a Convolutional Neural Network.” [Online]. Available: <https://www.tensorflow.org/tutorials/layers>. [Accessed: 20-Feb-2018].
- [20] Google, “TensorFlow.” [Online]. Available: <https://www.tensorflow.org/>. [Accessed: 01-Feb-2018].
- [21] T. Wang, Y. Chen, M. Zhang, J. I. E. Chen, and H. Snoussi, “Internal Transfer Learning for Improving Performance in Human Action Recognition for Small Datasets,” *IEEE Access*, vol. 5, pp. 17627–17633, 2017.
- [22] TensorFlow, “TensorBoard: Visualizing Learning,” 2018. [Online]. Available: https://www.tensorflow.org/programmers_guide/summaries_and_tensorboard. [Accessed: 15-Apr-2018].
- [23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, vol. 22, no. 4. MIT Press, 2016.
- [24] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 248–255, 2009.
- [25] Google, “TensorFlow for Poets.” [Online]. Available: https://ml-days-prd.s3.amazonaws.com/slides/speakers/slides/2/TensorFlow_for_Poets_1_.pdf. [Accessed: 01-Feb-2018].
- [26] “ML Cheatsheet - Loss Function.” [Online]. Available: http://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html. [Accessed: 10-Mar-2018].
- [27] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Amplitude-selective filtering for remote-PPG,” *Biomed. Opt. Express*, vol. 8, no. 3, pp. 1965–1980, 2017.
- [28] E. M. Tapia, S. S. Intille, W. Haskell, K. W. J. Larson, A. King, and R. Friedman, “Real-Time Recognition of Physical Activities and their Intensities Using Wireless Accelerometers and a Heart Monitor,” *Int. Symp. Wearable Comput.*, pp. 37–40, 2007.