Smart Health 14 (2019) 100082

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

ActiPPG: Using deep neural networks for activity recognition from wrist-worn photoplethysmography (PPG) sensors

Mehdi Boukhechba *, Lihua Cai, Congyu Wu, Laura E. Barnes

Department of Engineering Systems and Environment, University of Virginia, United States

ARTICLE INFO

Keywords:
Wearable sensing
Activity recognition
Smartwatches
Photoplethysmography (PPG)
Neural networks
Deep learning

ABSTRACT

Sensor-based activity recognition seeks to provide higher-level knowledge about human activities from multiple sensors such as accelerometer and gyroscope. Thanks to growing ubiquity of sensor-rich smartphones and wearable devices, activity recognition research has made tremendous progress in recent years. Many sensors, such as motion sensors and cameras, have been previously used for this task. In this paper, we present the first work to recognize human activities solely from photoplethysmography (PPG) data. PPG sensors use a light-based technology to sense the rate of blood flow as controlled by the heart's pumping action. Although they are not originally intended to sense body motions and gestures, in this paper, we propose a novel method to extract meaningful features from the PPG to predict human activities. We exploit PPG signals' susceptibility to corruption by noise introduced by motion artifact and extract, as opposed to discard, the motion artifact signals combined with cardiac and respiratory signals to predict the types of activities performed by users. We combine convolutional and recurrent neural networks to predict different types of daily activities (e.g., walking, running, jumping) from the raw PPG signal. Results using data generated by wrist-worn smartwatches of 12 participants demonstrate the feasibility of our approach at predicting five types of activities (standing, walking, jogging, jumping, and sitting), and highlight new insights about how to use PPG sensors for human activity recognition. Results also highlight the importance of extracting motion artifact signals to understand human behaviors.

1. Introduction

Inferring and monitoring human activities using sensor embedded devices (e.g., smart home appliances and mobile devices) have been used in a wide range of applications including but are not limited to security and surveillance (Chen, Wei, & Ferryman, 2013; Taha, Zayed, Khalifa, & El-Horbaty, 2015), smart home (Cicirelli et al., 2016), health care (Boukhechba, Chow, Fua, Teachman, & Barnes, 2018, Boukhechba et al., 2018), and human computer interaction (Han, Shao, Xu, & Shotton, 2013). Human activity recognition (HAR) leverages a variety of sensors, either stand-alone or embedded, to collect raw signals describing the surrounding environment. HAR applies machine learning models to recognize the underlying activity using the features extracted from the raw sensor traces. Among the most utilized sensors are accelerometer (Attal et al., 2015), GPS (Boukhechba, Bouzouane, Bouchard, Gouin-Vallerand, & Giroux, 2015), camera or image-based sensor (Chen et al., 2013). There are different strengths and weaknesses in each of these sensing modalities. For example, tri-axis accelerometer consumes little energy and is robust in collecting motion signals when placed on the user's body, but it is

* Corresponding author.
E-mail addresses: mob3f@virginia.edu (M. Boukhechba), lc3cp@virginia.edu (L. Cai), cw9dd@virginia.edu (C. Wu), lb3dp@virginia.edu (L.E. Barnes).

https://doi.org/10.1016/j.smhl.2019.100082

Available online 2 November 2019
2352-6483/

sensitive to position and orientation. Location trajectories collected by GPS are able to reveal location-associated activities (e.g., shopping at a mall), but when obstructed by buildings in dense urban areas, GPS may not work properly. Image-based sensor, on the other hand, captures directly the nature of activities, but it creates significant privacy concerns. Since no single sensor is ideal for all application scenarios, researchers have proposed combining multiple sensor modalities and leveraging new sensing modalities such as WiFi and Bluetooth (Wang et al., 2015, 2017) to improve the prediction performance.

The formality in which the sensor is designed is as important as the choice of sensing modality in HAR. For example, some earlier HAR works applied a set of accelerometers that were placed on different positions on the human body (Attal et al., 2015; Siewiorek et al., 2003). The placement and quantity of accelerometers, as a result, became key determinants in recognition performance. These approaches create at least two challenges. First, users will need to pay for these additional sensing systems if they want to enjoy the benefits brought by HAR; second, wearing or carrying these sensors is intrusive and unrealistic, and thus is hard to scale. Fortunately, the rapid advancements in sensing technologies lowered the cost of sensor-rich smartphones and wearable devices, and led to the pervasiveness and ubiquity of these mobile devices. Consequently, more recent HAR works (Bhattacharya & Lane, 2016; Jiang & Yin, 2015; Radu et al., 2016; Ronao & Cho, 2016) have been conducted using smartphone and smartwatches.

PPG sensor is among the most popular sensors embedded in today's smartwatches and wristbands. Photoplethysmography (PPG) signals are optically obtained plethysmograms that can be used to detect blood volume changes in the microvascular bed of tissues (Joseph, Joseph, Titus, Thomas, & Jose, 2014). In other terms, it monitors heart rate (HR) by shining light into the body and measuring the amount of light that is reflected back to estimate the blood flow. Unlike the electrocardiography (ECG) monitoring that requires placements of sticky metal electrodes on the body skin in order to monitor electrical activity from heart and muscles (Biagetti, Crippa, Falaschetti, Orcioni, & Turchetti, 2017), PPG monitoring can be performed at peripheral sites on the body, requiring less intrusive body contact. As a result, PPG sensors are used more frequently for heart rate monitoring in personal fitness devices such as smartwatches and wristbands.

One major limitation of PPG sensors is their sensitivity to motion artifacts (MAs). Accurate estimation of PPG signal recorded from subject's wrist, when the subject is performing various physical activities (e.g., walking and upper body movements), is often a challenging problem as the raw PPG signal is severely corrupted by MAs (Keerthiveena & Esakkirajan, 2017). These are principally due to the relative movement between the PPG light source/detector and the wrist skin of the subject during motion (Joseph et al., 2014). In this paper, we propose to exploit the MAs to recognize human activities. Given the PPG signal becomes corrupted during physical activities, we propose to extract the noise detected during those activities and convert it into features to predict the type of physical activities performed by users.

We present a novel approach that uses end-to-end neural networks for HAR from PPG signals. Our model takes as input preprocessed PPG signals, and output the learned activity labels without manual feature extraction. PPG sensors have not been applied in HAR research because they are not designed to capture motion signals, just like WiFi and Bluetooth signals are not initially intended for HAR either. Using PPG sensors for HAR has several advantages. It adds no additional cost to users with one of these PPG embedded smartwatches or wristbands; such wearable devices are becoming ubiquitous, which enables scale-up of these HAR systems to benefit the general public; and lastly, it can either be used alone when other HAR sensors are unavailable, or combined with them to augment recognition performance.

The remainder of the paper is structured as follows. We describe related works of HAR and PPG signal processing in Section 2 and provide background information about PPG sensors in Section 3. Then, in Section 4 we discuss our approach in detail and in Section 5 we present our performance evaluation and results. Finally, we discuss the practicality and limitations of our current work in Section 6 and make our concluding remarks in Section 7.

2. Related works

In this section, we first summarize the current status in HAR in terms of the types of activities being recognized, the sensors applied to achieve the recognition tasks, and the state-of-the-art of machine learning techniques adopted in existing works. We also provide a summary on the applications of PPG sensor in various tasks, and the existing preprocessing techniques to prepare the raw PPG signal for analyses.

2.1. Activity recognition

Human activities can be broadly categorized into ambulatory (e.g., walking and sitting), transportation mode (e.g., in a vehicle and riding a bike), phone usage (e.g., making a call), daily activities (e.g., eating and watching TV), and many more (Incel, Kose, & Ersoy, 2013; Lara & Labrador, 2013). However, there is no rigid demarcation to separate one activity of a certain category from one in another category. In essence, the targeted activities are entirely application driven. For example, in order to promote healthy living style and prevent obesity, step counts, distance traveled, and physical inactivity (e.g., sitting and being still, or the lack of movement) can be targeted (Incel et al., 2013). In our current work, we collect data using smartwatches to recognize ambulatory activities including standing, walking, jogging, jumping, and sitting.

Different sensors have different aptitudes to recognize different activities. Therefore, choice of sensors is based on the types of activities we aim to recognize. Effectiveness to recognize the targeted activities, costs, form factor, and intrusiveness are all important factors in the decision making. Accelerometer is the most explored sensor in recognizing ambulatory activities because of its effectiveness at monitoring actions that involve repetitive body motions, such as walking, running, cycling, and climbing stairs (Shweta et al., 2017). Gyroscope and magnetometer are also frequently used together with accelerometer (Ha & Choi, 2016) to improve the recognition performance. Other sensors such as GPS, light and audio sensors (Boukhechba et al., 2015; Chen, Hoey, Nugent, Cook, & Yu, 2012) have been used to add additional contextual features (e.g., current location, acoustic environment) to further improve recognition performance. More recently, some researchers proposed the use of wireless signals such as WiFi to recognize indoor ambulatory activities (Wang et al., 2015, 2017). This is made possible because human activities cause perturbation in the WiFi signals emitted and received by nearby WiFi devices. The advantages of using WiFi signals for HAR include being device-free, better indoor coverage, and privacy protection. (Wang et al., 2017).

In our current work, we instead propose to leverage a new sensing modality, using PPG sensor embedded in smartwatches and wristbands to recognize ambulatory activities. PPG sensor is mostly used to monitor cardiac and respiratory activities. The motion artifacts or MAs are usually removed from the raw PPG signal before it can be used to accurately measure heartbeats and respiration cycles. We take an opposite approach when applying PPG signal for HAR. Rather than discarding the MAs, we hypothesize that it can provide predictive power in recognizing different ambulatory activities. We apply a signal preprocessing procedure that separate the raw PPG signal into three signals (cardiac, respiration, and motion artifact), and feed them into end-to-end neural network models to predict five different ambulatory activities.

Once the sensors are chosen to recognize certain targeted activities, the natural next step is to determine which analytical approach for the recognition tasks. Specifically, various machine learning methods (e.g., random forest, decision tree, hidden Markov Model, and deep learning) have been explored in many existing HAR works (Attal et al., 2015; Casale, Pujol, & Radeva, 2011; Jiang & Yin, 2015; Lu et al., 2017; Ronao & Cho, 2016; Yang, Nguyen, San, Li, & Krishnaswamy, 2015). These methods can be categorized based on how they conduct featurizations (e.g., manually-crafted features versus automatic feature extractions). The former category corresponds to methods that are non-deep learning based, while the later category corresponds to deep learning approaches (e.g., feed-forward neural networks, convolutional neural networks, recurrent neural networks, and certain combinations of them).

In non-deep learning based approaches, Pierluigi et al. proposed a wearable system that placed accelerometer on user's chest, and applied random forest models on 319 manually extracted features from accelerometer signal to recognize walking, climbing stairs, talking with a person, staying standing, and working at computer (Casale et al., 2011). They achieved accuracies of 90% or above. Lu et al. applied an unsupervised clustering method called Molecular Complex Detection or MCODE to recognize physical and sporting activities using 19 features extracted from smartphone embedded accelerometers that were placed on study volunteers' waist (Lu et al., 2017). The MCODE clustering method performed better than most other clustering algorithms such as GMM and K-means++ with accuracy and F-score of 0.88 and 0.85, respectively. In (Attal et al., 2015), different classification techniques, including both supervised (e.g., K-Nearest Neighbors, SVM, random forest), and unsupervised (e.g., Gaussian Mixture Models, K-Means, Markov Chain and Hidden Markov Models) are reviewed in the context of recognizing daily living activities of elderly people with three inertial sensors being placed at various points of upper/lower body limbs, and both time-domain and frequency domain features. This study identified K-NN and HMM as the best algorithms in supervised and unsupervised methods, respectively. One major challenge that arose from these existing works is that they were all conducted in very distinctive settings in terms of placements and quantity of sensors, targeted activities, and featurization. As noted by (Jordao, Nazare, Sena, & Schwartz, 2018), this makes it impossible to draw definitive insights in how to optimize HAR. In an ideal scenario, a HAR system should not add any extra burden to their users (i.e., requiring the user to carry a new device for the sole purpose of HAR), exhibit different degrees of functionality contingent upon placement (i.e., only functional in certain fixed position), or require significant efforts in featurizations, which greatly reduce its generalizability across different HAR tasks.

Recent years have witnessed fast advancement of deep learning, which achieves unparalleled performance in many areas such as visual object recognition, natural language processing, and logic reasoning (Nweke, Teh, Al-garadi, & Alo, 2018). In fact, deep learning has been applied in image-based HAR, which applies sensors such as camera to capture both still images and video for activity recognition (Chen et al., 2012). Different from traditional predictive methods, deep learning can greatly reduce the effort of designing features, and learn much more high-level and abstract features by training an end-to-end neural network. In addition, with the deep network structure it is more feasible to perform unsupervised and incremental learning. Therefore, deep learning is becoming an ideal approach for HAR and has been widely explored in recent works (Liu et al., 2018). We applied a combination of convolutional neural networks and recurrent neural networks in an end-to-end design for our activity recognition task. While our current work does not focus on proposing new deep learning algorithms for HAR, interested readers can refer to the following HAR studies (Alsheikh et al., 2016; Bhattacharya & Lane, 2016; Cho & Yoon, 2018; Jiang & Yin, 2015; Radu et al., 2016; Ronao & Cho, 2016; Wang, Chen, Hao, Peng, & Hu, 2019; Yang et al., 2015), for more details about theory and implementation of deep network models.

2.2. Motion artifacts in PPG signals

The existence of Motion Artifacts (MAs) in raw PPG signal constitutes a major challenge for its application in accurately measuring heartbeats and respiration cycles. Various methods have been proposed to remove MAs from raw PPG signal. However, these denoising methods may be leveraged to serve our purpose, that is to separate MAs from raw PPG signal and use it in HAR. We summarize several most important works in MAs removal below. This may ultimately help readers interested in replicating our study explore MA extraction methods other than ours.

Chan et al. proposed to apply a LMS filter with automatic step-size control to mitigate the effects of MAs in PPG recordings for long-term patient monitoring (Chan & Zhang, 2002). However, this approach requires a reference sensor that detects a noise component that is correlated with the motion artifacts in the PPG signal. Since we did not have a reference sensor set up to collect the PPG data, this approach is not applicable to our setting. Ram et al. applied an adaptive step-size least mean squares (AS-LMS) adaptive filter for reducing MA in corrupted PPG signals (Ram, Madhav, Krishna, Komalla, & Reddy, 2012). Without needing a reference noise sensor, synthetic noise signal is generated internally from the MA corrupted PPG signal. Kim et al. exploited the quasi-periodicity of the PPG signal and the independence between the PPG and the MAs to remove MAs from raw PPG signals (Kim & Yoo, 2006). A preprocessor, which consists of period detection, block interleaving, low-pass filtering, and block de-interleaving, is applied to recognize the periodicity of heart pulsation to enhance the PPG signals. After this preprocessing step, Independent Component Analysis (ICA) is applied to separate the MAs from the PPG signals. However, this approach is applicable when multiple PPG signals are available. In our current work, instead of using ICA and/or the proposed adaptive filter in the above works, we combine Fast Fourier Transform (FFT) and Butterworth bandpass filters to extract different components from raw PPG signals, as described in (Daud & Sudirman, 2015). More details are provided in Section 4.2.

To the best of our knowledge, this is the first work attempting to recognize physical activities solely from PPG signals. In previous work, PPG has been fused with accelerometer signals for HAR (Biagetti et al., 2017), but not used independently. Using raw PPG alone performed poorly (accuracy around 44.7%) at recognizing four physical activity levels (walking, running and two resistance levels of biking: low and high). However, when combined with accelerometer, authors show a significant improvement reaching an accuracy of 78%. In this paper, we show that our analytical approach is able to accurately recognize human activities using PPG signals alone. This is mainly due to our innovative feature extraction method that extracts comprehensive features descriptive of cardiac responses, respiration rates, and motion artifacts; and due to our deep prediction model that is able to efficiently learn a representation of PPG signal for each type of physical activity.

3. Background

A typical PPG device contains a light source and a photodetector. The light source emits light to a tissue and the photodetector measures the reflected light from the tissue. The reflected light is proportional to blood volume variations. Similar to ECGs, PPG waves can also help to diagnose cardiac arrhythmias (irregular heartbeat) because they reliably manifest cardiac and respiratory activities (Zhao et al., 2017). Most common PPG sensors use an infrared light emitting diode (IRLED) or a green LED as the main light source. Although there are other LED sensors with different colors to measure hemoglobin, green LED is considered the most commonly used. This is simply because it penetrates more deeply into tissue and therefore can provide measurements that are more accurate. PPG sensors also use a photodetector to measure the intensity of reflected light from the tissue. The blood volume changes can then be measured (calculated) based on the amount of the detected light (see Fig. 1).

HR monitoring techniques that rely on PPG sensors have several advantages over traditional ECG-based systems. For instance, PPG sensors use simpler hardware implementation and have lower costs, and for operation, only a single sensor is required to be placed on the body. This is in contrast with traditional ECG recordings. A traditional ECG-based system requires at least three bio-electrodes placed on different body locations (e.g., right arm, left arm, and right leg) to be able to operate effectively. This requirement greatly restricts the patients' flexibility of motion (Zhao et al., 2017).

Several factors can affect PPG recordings. These factors are sensing, biological, and cardiovascular factors. Tissue modifications generated by voluntary or involuntary movements can create alterations of inner tissues, such as muscle movement and dilation of tissues. The receiving light will be modified due to these movements, generating a different signal called corrupted signal due to motion artifacts (Zhao et al., 2017). The goal of this paper is to show how to use the noise extracted from PPG signals to predict the type of motion performed by the user.

4. Our approach

In this section, we describe our study design, preprocessing technique, and modeling approach for HAR using PPG data.

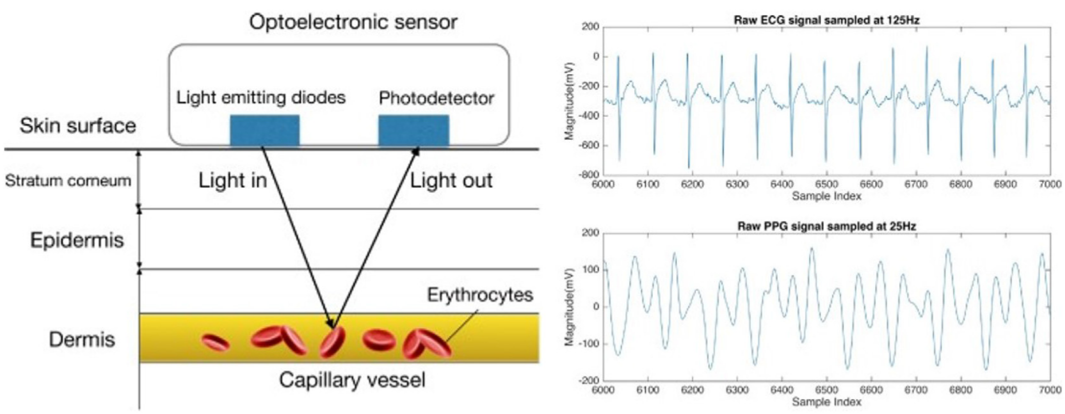

Fig. 1. Working principle of PPG sensors and examples of ECG and PPG signals (Zhao, Sun, Wan, & Wang, 2017).

4.1. Study design

12 participants (6 females) were recruited to participate in a controlled activity recognition experiment. Participants were asked to perform 5 activities (standing still, walking, jogging, jumping, and sitting) while wearing a Huawei Watch 2 smartwatch. A custom Android app was previously installed on the smartwatches to collect 100HZ PPG data and sync it periodically to a secure Amazon web server. As shown in Fig. 2 participants were asked to walk and jog naturally following a predefined path. Standing, sitting and jumping where performed where all performed on the same place (see the path start on Fig. 2). Each activity lasted around 90 s with around 1 min gap between each two consecutive activities. The participants were asked to perform this sequence of activities three times which resulted to around 16000 s of labeled PPG signals. The exact start and end time of each activity were recorded by a research assistant. Data was labeled after the end of the experiment to match each second of PPG signal to the appropriate physical activity.

4.2. Signal preprocessing

The frequency spectrum of the MA-corrupted PPG signal consists of various frequency components. The pulsatile PPG cardiac portion (0.5–4 Hz), respiratory activity (0.2–0.35 Hz), and MA noise component (0.1 Hz or more) generated by the motion artifact (Ban & Kwon, 2016). Accordingly, we transformed the PPG signal using FFT and decomposed it to three signals: (1) cardiac signal (CS), (2) respiration signal (RS), and (3) motion artifact signal (MS) using a 5th order bandpass Butterworth filter (Daud & Sudirman, 2015). This filter was applied three times to extract each type of signal with its corresponding frequency boundaries (e.g., [0.5-4HZ] to create RS). Let P represent the preprocessed signal. The preprocessing step can be defined as:

$$P = f_{bp}(P_o, c_l, c_u), \tag{1}$$

where c_l and c_u represent the lower and upper cutoff frequencies of the bandpass filter f_{bp}. P_o is the original PPG signal.

Consequently, the initial PPG signal is replaced by three new signals describing cardiac, respiration, and motion artifact levels. We augment our feature space by adding these three dimensions (CS, RS, and MS). This process is demonstrated by the example presented in Fig. 3 that compares the raw PPG signal with the extracted CS, RS, and MS in different activities. We can observe how the raw PPG signal is not visually indicative of the type of activity performed by the subject. On the contrary, we can see how the intensity of the physical activity impacted the CS, RS, and MS differently. The figure also shows how the motion artifact signal is more pronounced in jogging when compared to standing still. This comparison will be further studied in the experimentation section to provide tangible metrics as to why this preprocessing step is important.

4.3. End-to-end deep neural network modeling

Recent studies in HAR have shown limited generalizability of hand-crafted features across sensors and HAR applications (Sun et al., 2018). On the contrary, automatic feature representations using deep learning approaches enable extraction of predictive features that can be generalized across different HAR applications without reliance on domain expert knowledge. They provide the ability to learn features from raw sensor data with minimal efforts in feature engineering while achieving equal or better recognition performance

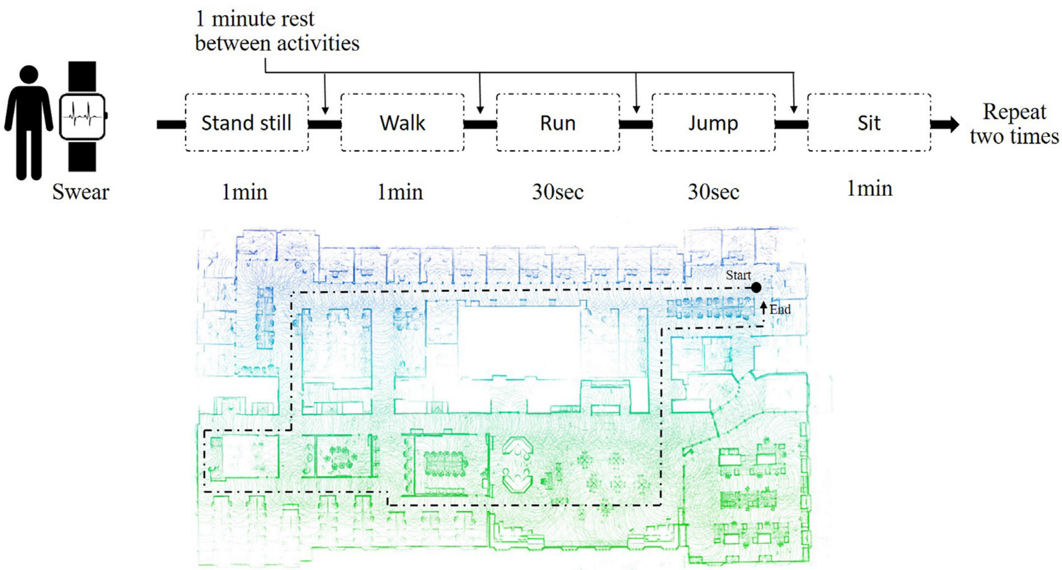

Fig. 2. Design of our data collection experiment. Swear is the name of the Android Watch app used to collect data.

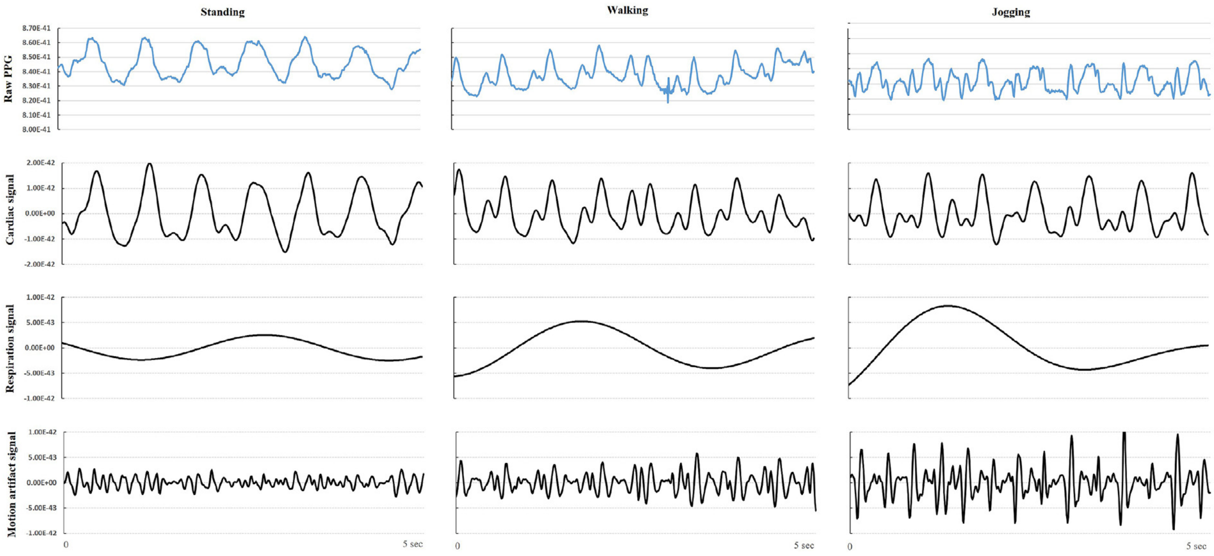

Fig. 3. Example of 5 s recordings of the raw PPG data (top of the figure). The three signals in black describe the cardiac, respiration, and motion artifact signals obtained after preprocessing. The variations of those signals during three types of activities: standing, walking, and jogging are displayed from left to right respectively.

(Nweke et al., 2018). Using multiple layers of abstraction, deep learning methods learn intricate feature representations from raw sensor data and discover the best pattern to improve recognition performance.

Our approach is inspired by the work of Ordóñez and Roggen (Ordóñez & Roggen, 2016) that proposes a framework for using end-to-end deep neural networks for multimodal HAR. Our model combines convolutional and recurrent layers. The convolutional layers act as feature extractors and provide abstract representations of the three CS, RS, and MS data in feature maps. The recurrent layers model the temporal dynamics of the activation of the feature maps. Fig. 4 describes our overall modeling approach. First, the three signals CS, RS, and MS are segmented using a 2-s sliding window with 50% overlap. This resulted to 16165 2-s windows used to train and test our models. Segmented signals are then transformed through four convolutional operations. Convolutional layers process the input only along the axis representing time. The number of dimensions is the same for every feature map in all layers. These convolutional layers employ rectified linear units (ReLUs) to compute the feature maps. Layers 6 and 7 are recurrent dense layers. The choice of the number of recurrent layers is made following the results presented in (Karpathy, Johnson, & Li, 2015), where the authors showed that a depth of at least two recurrent layers is beneficial when processing sequential data. Recurrent dense layers adapt their internal state after each time step. Here, the inputs of Layer 6 at time t are the elements of all of the feature maps at Layer 5. The activation of the recurrent units is computed using the hyperbolic tangent function. The output of the model is obtained from a softmax layer (a dense layer with a softmax activation function), yielding a class probability distribution for every single time window. The shorthand description of this model is: C(64)-C(64)-C(64)-C(64)-R(128)-R(128)-Sm.

Models are trained in a fully-supervised way, backpropagating the gradients from the softmax layer through to the convolutional layers. The network parameters are optimized by minimizing the cross-entropy loss function using mini-batch gradient descent with the RMSProp update rule (Dauphin, de Vries, Chung, & Bengio, 2015). After experimenting with multiple per-parameter learning rate updates, we found that RMSProp consistently offered the best results with the widest tolerance to the learning rate setting. For the sake of efficiency, when training and testing, data are segmented on mini-batches of a size of 100 data segments. Using this configuration, an accumulated gradient for the parameters is computed after every mini-batch. The model is trained with a learning rate of $10e^3$ and a

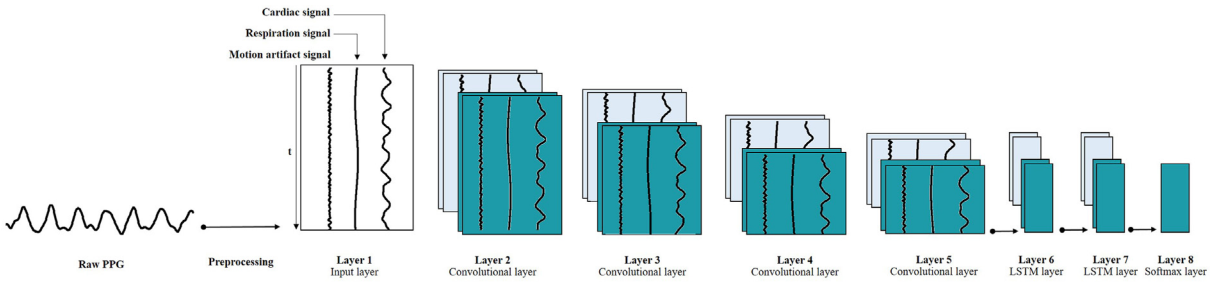

Fig. 4. Our End-to-end deep neural network model.

decay factor of = 0.9. Weights are randomly orthogonally initialized. We introduce a drop-out operator on the inputs of every dense layer, as a form of regularization. This operator sets the activation of randomly-selected units during training to zero with probability $p = 0.5$.

5. Results

Results are presented in Table 1, comparing our model with six baseline models as follows: BM_1, BM_2, BM_3, BM_4 are models that use the raw PPG, cardiac signal, respiration signal, and motion artifact signal respectively. The goal of these baseline models is to investigate the benefit of the reprocessing step. It also allows to compare the importance of each component of the PPG signal (cardiac, respiration and motion artifact signals) at predicting the performed activities. We also compare our method with BM_5 and BM_6 that use similar features to our model (combining CS, RS, MS) but with a random forests instead of an end-to-end neural network. In BM_5 we feed the raw three signals to random forests, and the goal is to investigate the benefit of using a neural network architecture. In BM_6 we use a random forests with the following time and frequency domains hand-crafted features: 25, 50, and 75 percentile, mean, min, max, standard deviation, permutation entropy and the log energy of each of three CS, RS, and MS signals. Those features are commonly used to featurize sensor signals for HAR (Dehzangi and Sahu, 2018). The goal of including this baseline model is to investigate the benefit of our CNN-based feature extraction technique.

The evaluation of each model is performed using leave-one-subject-out cross-validation. The F1-score for each type of activity is displayed. Results presented in Table 1 show that preprocessing the PPG signal and combining CS, RS, and MS lead to better performance than using the raw PPG signal (See our model vs. BM_1). Moreover, combing CS,RS, and MS resulted in a better performance than using each signal separately (See our models vs. BM_2, BM_3, and BM_4).

When comparing BM_2, BM_3, and BM_4, results suggest that the motion artifact signal is a better predictor than the cardiac and respiration signal. This suggests that the predictability of activities from PPG is not resulted from the elevation of heart rate and respiration rates only, but also the motion artifacts generated when performing those activities. This also speaks to the importance of extracting motion artifact signals to understand human behaviors, a step that has been neglected by the literature. Our model also outperformed BM_5 and BM_6. This suggests that using our end-to-end neural network architecture yielded a better performance than traditional models with hand-crafted features. It also shows the importance of our recurrent neural network architecture that captures the time dependencies during the prediction.

Moreover, we can also observe that sitting activity had the worst F1-score across the different models, this suggests that those models were not able to accurately differentiate between sitting and other activities using PPG signals. To further investigate this finding, we present in Table 2 the confusion matrix of our model.

The confusion matrix suggests that our model was not able to accurately differentiate between sitting and standing and between jogging and jumping. This may be explained by the fact that those activities are similar in terms of cardiac responses, respiration rates, and motion artifacts. For instance, in both standing and sitting activities, the participants were in an idle state (not moving their arms). This is expected to generate very similar PPG signals characterized by low cardiac response, low respiration activity, and low rate of motion artifacts.

In Table 3, we explore the effect of merging similar activities together on the performance of our model. Results show that our models had an F1-score of 0.92 when predicting if the participants were moving versus not moving. Also, our model had an F1-score of 0.86 when predicting if the participants are in a stationary state, walking, or jogging/jumping. This means that our model had a higher performance when predicting less granular activities. In other terms, it seems that our PPG-based HAR model is able to accurately differentiate between activities with a wide difference in the motion range (e.g., walking and jogging) and not activities with similar motion intensities (e.g., standing and sitting).

6. Discussion

The goal of this work is not to replace the widely used motion sensors, but to demonstrate how PPG signals can be leveraged in activity recognition. For example, the proposed approach could be used to infer human activities when accelerometer or other sensors are not available. PPG signals could also be fused with other sensors to improve the prediction accuracy. Moreover, PPG signals could be

Table 1
Results comparison between our model and six baseline models. Overall F1-score is displayed alongside F1-score for each activity (standing, walking, jogging, jumping, and sitting).

Model	overall F1	Standing	Walking	Jogging	Jumping	Sitting
Sample size	16165	3521	3907	3221	2193	3323
Our model	**0.78**	**0.85**	**0.81**	**0.78**	**0.75**	**0.59**
BM_1: **Raw PPG**	0.59	0.68	0.62	0.46	0.39	0.38
BM_2: **Cardiac Signal (CS)**	0.57	0.61	0.55	0.43	0.44	0.40
BM_3: **Respiration Signal (RS)**	0.26	0.27	0.23	0.22	0.27	0.20
BM_4: **Motion Artifact Signal (MS)**	0.62	0.63	0.67	0.59	0.55	0.51
BM_5: **Random forests with raw CS+RS+MS**	0.70	0.81	0.75	0.73	0.65	0.52
BM_6: **Random forests with featurized CS+RS+MS**	0.69	0.71	0.69	0.69	0.59	0.54

Table 2
Confusion matrix of our model. Cells are displayed in percentage.

	Standing	Walking	Jogging	Jumping	Sitting
Standing	**0.78**	0.06	0.02	0.02	**0.12**
Walking	0.04	**0.72**	0.10	0.09	0.05
Jogging	0.03	0.09	**0.65**	**0.20**	0.03
Jumping	0.05	0.10	**0.21**	**0.62**	0.03
Sitting	**0.31**	0.05	0.03	0.03	**0.51**

Table 3
Performance evaluation when merging activities together.

	Moving vs. Idle	Idle vs. Walking vs. Jogging/Jumping
F1-score	0.92	0.86
SD	0.03	0.03

used to save battery power by turning off the motion sensors if they are only intended for HAR. In this case, the same PPG signal used to calculate heart rate will be used to extract meaningful information about the performed activities. By using the same sensor for multiple purposes, researchers may be able to reduce resource utilization for efficient sensing.

The proposed method can operate online, as our preprocessing step is usually already performed on devices to estimate HR. In other words, software running on smartwatches would already have been applying FFT to extract the cardiac signal in order to estimate HR, which implies that our method will not require any extra preprocessing computation. After preprocessing the signal, the next step would be to feed the result of FFT to a pre-trained neural network model, which is considered to be computationally inexpensive (Pertusa, Gallego, & Bernabeu, 2018).

This paper has several limitations. First, the experiment was conducted with a small sample size in a semi-controlled environment, which may limit generalizability of our findings. Second, we did not compare our method with other sensors such as accelerometer and gyroscope. The reason is that it's already proven in existing literature that those motion sensors, which are designed for this kind of tasks, are performing very well (Bhattacharya & Lane, 2016; Jiang & Yin, 2015; Radu et al., 2016; Ronao & Cho, 2016). We focused our paper on the unexplored usability of PPG for HAR, and investigated how to use the different components of PPG to predict human activities. Despite those limitations, the current research opens new perspectives on how to use other sensing modalities than motion sensors to estimate human activities.

7. Conclusions

In this paper we study the feasibility of recognizing human activities from PPG sensors. The proposed method exploits a characteristic of PPG signals that get corrupted when a significant motion is performed. We propose to extract the noise detected in the corrupted signal and use it with other cardiac and respiration signals to predict the type of motion performed by the user. Our methodology decomposes the PPG signal into three cardiac, respiration, and motion artifact signals, and uses an end-to-end deep neural network that combines convolutional and recurrent layers to predict human activities. The proposed method has been tested using data generated by 12 participants in a controlled experiment. Results provide preliminary evidences of the effectiveness of this method at predicting several granularity levels of human activities (e.g., moving vs. idle, idle vs. walking vs. jogging/jumping). Our results also highlight the importance of extracting motion artifact signals and how to use them to understand human behaviors. Future work will involve studying the effect of combining PPG data with other low-energy sensors such as accelerometer and gyroscope for HAR.

Declaration of Competing Interest

The authors declare no conflict of interest.

References

Alsheikh, M. A., Selim, A., Niyato, D., Doyle, L., Lin, S., & Tan, H.-P. (2016). Deep activity recognition models with triaxial accelerometers. In *Workshops at the thirtieth AAAI conference on artificial intelligence*.
Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., & Amirat, Y. (2015). Physical human activity recognition using wearable sensors. *Sensors, 15*(12), 31314–31338.
Ban, D., & Kwon, S. (2016). Movement noise cancellation in PPG signals. In *2016 IEEE international conference on consumer electronics (ICCE)*. IEEE. https://doi.org/10.1109/icce.2016.7430517.
Bhattacharya, S., & Lane, N. D. (2016). From smart to deep: Robust activity recognition on smartwatches using deep learning. In *2016 IEEE international conference on pervasive computing and communication workshops (PerCom workshops)* (pp. 1–6). IEEE.
Biagetti, G., Crippa, P., Falaschetti, L., Orcioni, S., & Turchetti, C. (2017). Human activity recognition using accelerometer and photoplethysmographic signals. In *International conference on intelligent decision technologies* (pp. 53–62). Springer.

Boukhechba, M., Bouzouane, A., Bouchard, B., Gouin-Vallerand, C., & Giroux, S. (2015). Online recognition of people's activities from raw gps data: Semantic trajectory data analysis. In *Proceedings of the 8th ACM international conference on PErvasive technologies related to assistive environments, PETRA '15*. New York, NY, USA: ACM. https://doi.org/10.1145/2769493.2769498, 40:1–40:8.

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. https://doi.org/10.2196/10101.

Boukhechba, M., Daros, A. R., Fua, K., Chow, P. I., Teachman, B. A., & Barnes, L. E. (2018). *Smart Health, 9–10*, 192–203. https://doi.org/10.1016/j.smhl.2018.07.005.

Casale, P., Pujol, O., & Radeva, P. (2011). Human activity recognition from accelerometer data using a wearable device. In *Iberian conference on pattern recognition and image Analysis* (pp. 289–296). Springer.

Chan, K., & Zhang, Y. (2002). Adaptive reduction of motion artifact from photoplethysmographic recordings using a variable step-size lms filter. In *Sensors, 2002. Proceedings of IEEE* (Vol. 2, pp. 1343–1346). IEEE.

Chen, L., Hoey, J., Nugent, C. D., Cook, D. J., & Yu, Z. (2012). Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42*(6), 790–808. https://doi.org/10.1109/tsmcc.2012.2198883.

Chen, L., Wei, H., & Ferryman, J. (2013). A survey of human motion analysis using depth imagery. *Pattern Recognition Letters, 34*(15), 1995–2006.

Cho, H., & Yoon, S. (2018). Divide and conquer-based 1d cnn human activity recognition using test data sharpening. *Sensors, 18*(4), 1055.

Cicirelli, F., Fortino, G., Giordano, A., Guerrieri, A., Spezzano, G., & Vinci, A. (2016). On the design of smart homes: A framework for activity recognition in home environment. *Journal of Medical Systems, 40*(9), 200.

Daud, S., & Sudirman, R. (2015). Butterworth bandpass and stationary wavelet transform filter comparison for electroencephalography signal. In *2015 6th international conference on intelligent systems, modelling and simulation*. IEEE. https://doi.org/10.1109/isms.2015.29.

Dauphin, Y., de Vries, H., Chung, J., & Bengio, Y. (2015). Rmsprop and equilibrated adaptive learning rates for non-convex optimization. In *NIPS*.

Dehzangi, O., & Sahu, V. (2018). In *2018 24th international conference on pattern recognition (ICPR)*. IEEE. https://doi.org/10.1109/icpr.2018.8546311.

Ha, S., & Choi, S. (2016). Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In *2016 international joint conference on neural networks (IJCNN)*. IEEE. https://doi.org/10.1109/ijcnn.2016.7727224.

Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced computer vision with microsoft kinect sensor: A review. *IEEE transactions on cybernetics, 43*(5), 1318–1334.

Incel, O. D., Kose, M., & Ersoy, C. (2013). A review and taxonomy of activity recognition on mobile phones. *BioNanoScience, 3*(2), 145–171.

Jiang, W., & Yin, Z. (2015). Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1307–1310). Acm.

Jordao, A., Nazare, A. C., Jr., Sena, J., & Schwartz, W. R. (2018). *Human activity recognition based on wearable sensor data: A standardization of the state-of-the-art*. arXiv preprint arXiv:1806.05226.

Joseph, G., Joseph, A., Titus, G., Thomas, R. M., & Jose, D. (2014). Photoplethysmogram (PPG) signal analysis and wavelet de-noising. In *2014 annual international conference on emerging research areas: Magnetics, machines and drives (AICERA/iCMMD)*. IEEE. https://doi.org/10.1109/aicera.2014.6908199.

Karpathy, A., Johnson, J., & Li, F. (2015). Visualizing and understanding recurrent networks. *CoRR*. abs/1506.02078arXiv:1506.02078.

Keerthiveena, B., & Esakkirajan, S. (2017). Denoising of PPG signal by wavelet packet transform. In *2017 international conference on intelligent computing, instrumentation and control technologies (ICICICT)*. IEEE. https://doi.org/10.1109/icicict1.2017.8342632.

Kim, B. S., & Yoo, S. K. (2006). Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Transactions on Biomedical Engineering, 53*(3), 566–568.

Lara, O. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials, 15*(3), 1192–1209.

Liu, Q., Zhou, Z., Shakya, S. R., Uduthalapally, P., Qiao, M., & Sung, A. H. (2018). Smartphone sensor-based activity recognition by using machine learning and deep learning algorithms. *International Journal of Machine Learning and Computing, 8*(2), 121–126. https://doi.org/10.18178/ijmlc.2018.8.2.674.

Lu, Y., Wei, Y., Liu, L., Zhong, J., Sun, L., & Liu, Y. (2017). Towards unsupervised physical activity recognition using smartphone accelerometers. *Multimedia Tools and Applications, 76*(8), 10701–10719.

Nweke, H. F., Teh, Y. W., Al-garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications, 105*, 233–261. https://doi.org/10.1016/j.eswa.2018.03.056.

Ordóñez, F., & Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors, 16*(1), 115. https://doi.org/10.3390/s16010115.

Pertusa, A., Gallego, A.-J., & Bernabeu, M. (2018). *Neurocomputing, 293*, 87–99. https://doi.org/10.1016/j.neucom.2018.03.005.

Radu, V., Lane, N. D., Bhattacharya, S., Mascolo, C., Marina, M. K., & Kawsar, F. (2016). Towards multimodal deep learning for activity recognition on mobile devices. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct* (pp. 185–188). ACM.

Ram, M. R., Madhav, K. V., Krishna, E. H., Komalla, N. R., & Reddy, K. A. (2012). A novel approach for motion artifact reduction in ppg signals based on as-lms adaptive filter. *IEEE Transactions on Instrumentation and Measurement, 61*(5), 1445–1457.

Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications, 59*, 235–244.

Shweta, Khandnor, P., & Kumar, N. (2017). A survey of activity recognition process using inertial sensors and smartphone sensors. In *2017 international conference on computing, communication and automation (ICCCA)*. IEEE. https://doi.org/10.1109/ccaa.2017.8229873.

Siewiorek, D., Smailagic, A., Furukawa, J., Krause, A., Moraveji, N., Reiger, K., et al. (2003). Sensay: A context-aware mobile phone. In *null* (p. 248). IEEE. http://www.computer.org/csdl/proceedings/iswc/2003/2034/00/20340248.pdf.

Sun, J., Fu, Y., Li, S., He, J., Xu, C., & Tan, L. (2018). Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. *Journal of Sensors*, 1–10. https://doi.org/10.1155/2018/8580959, 2018.

Taha, A., Zayed, H. H., Khalifa, M. E., & El-Horbaty, E.-S. M. (2015). Human activity recognition for surveillance applications. In *The 7th international conference on information technology*. Al-Zaytoonah University of Jordan. https://doi.org/10.15849/icit.2015.0103.

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters, 119*, 3–11.

Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. (2015). Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking* (pp. 65–76). ACM.

Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. (2017). Device-free human activity recognition using commercial wifi devices. *IEEE Journal on Selected Areas in Communications, 35*(5), 1118–1131.

Yang, J., Nguyen, M. N., San, P. P., Li, X. L., & Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. In *Twenty-fourth international joint conference on artificial intelligence*.

Zhao, D., Sun, Y., Wan, S., & Wang, F. (2017). SFST: A robust framework for heart rate monitoring from photoplethysmography signals during physical activities. *Biomedical Signal Processing and Control, 33*, 316–324. https://doi.org/10.1016/j.bspc.2016.12.005.