Wrist-worn Wearable Sensors to Understand Insides of the Human Body: Data Quality and Quantity

Pekka Siirtola
pekka.siirtola@oulu.fi
Faculty of Information Technology and Electrical Engineering
University of Oulu, Finland

Ella Peltonen
ella.peltonen@oulu.fi
Faculty of Information Technology and Electrical Engineering
University of Oulu, Finland

Heli Koskimäki
heli.koskimaki@oulu.fi
Faculty of Information Technology and Electrical Engineering
University of Oulu, Finland

Henna Mönttinen
henna.monttinen@oulu.fi
Faculty of Information Technology and Electrical Engineering
University of Oulu, Finland

Juha Röning
juha.roning@oulu.fi
Faculty of Information Technology and Electrical Engineering
University of Oulu, Finland

Susanna Pirttikangas
susanna.pirttikangas@oulu.fi
Faculty of Information Technology and Electrical Engineering
University of Oulu, Finland

ABSTRACT

Wearable sensors have become more commonly used in everyday basis and powerful in terms of computational capacity and sensing resources, including capability to collect data from different bio-signals. The data collected from everyday wearables offers huge opportunities to monitor people's everyday life without expensive laboratory measurements, including also behaviours and conditions only rarely seen in controlled laboratory environments. So far wearable sensors have mostly been used to monitor motion, but bio-sensor powered wearables can do a lot more: they can be used to monitor physiological reactions inside the human body as well as some psycho-physical reactions such as affection and stress. This development enables multiple interesting and important applications, such as early detection of diseases, seizures, and attacks. With stock wearables worn in everyday basis, one of the biggest challenges for such applications is the sensing data itself. In order to train reliable recognition and prediction models, high quality training data with labels needs to be collected. This paper focuses on lessons learned of challenges in data quality and quantity when such data sets are gathered. We discuss our own experiences when collecting data using wearable sensors for early detection of migraine attacks, but the same lessons learned can be generalized to other studies utilizing wearables for recognition medical symptoms and users' everyday behaviour.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Applied computing** → **Life and medical sciences**.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

KEYWORDS

wearable sensors, machine learning, lessons learned, bio-signals

ACM Reference Format:
Pekka Siirtola, Ella Peltonen, Heli Koskimäki, Henna Mönttinen, Juha Röning, and Susanna Pirttikangas. 2019. Wrist-worn Wearable Sensors to Understand Insides of the Human Body: Data Quality and Quantity. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The wearable sensor market is currently one of the most rapidly growing area in consumer electronics. Wearables provide users capability to monitor their everyday life from sleep to stress, but also researchers interesting opportunities to collect real-life data easily without heavy investments to expensive medical sensors. So far, wearables have most commonly used to monitor motion (i.e. so called activity and fitness trackers) through sensors such as accelerometer, gyroscope, and magnetometer. The progress of affordable sensor development has made it possible to measure also other features of human physiology in the wild. Especially wrist-worn wearable devices can include a wide range of sensors, including but not limited to, photoplethysmography (PPG), thermometer, and electrodermal activity sensor. These provide wider view to monitor human behaviour and bio-medical reactions. Sensor fusion, i.e. integration of information from multiple sensors, machine learning, and other data-driven analysis methods can thus be utilized to recognize even emotions and medical symptoms, and perform preliminary diagnosis.

One of the most interesting application field for wearable sensors is early detection of diseases, seizures, and attacks, as they can have a huge market potential and economic impact. Self-monitoring of symptoms by wearables offers opportunity of increasing number of patients be treated in early phase of their disease, to both reduce costs of treatments and possibly rise survival probabilities. Symptom recognition and prediction topics are thus become an active research area. For instance, heart rate variability can be used to predict epilectic attacks [4] and recognize sleep apnea [7]. Moreover, motion sensors (accelerometer, gyroscope, and magnetometer) can be used to detect Parkinson disease [8]. In addition, migraine attacks have been recognized beforehand in [9, 14] using fusion of different bio-signals. Indeed, it is estimated that 15% of people in developed countries suffer from migraine and the yearly costs of these in Europe are 111 Billion Euros [17]. By helping just a fraction of those people, large-scale savings can be achieved to society not to mention the improvement of quality of life of the actual patients.

Detection of emotions based on wrist-worn wearable sensors has also been studied, including amusement and stress [11]. In fact, when it comes to detecting emotions based on wearable sensors, detection of stress and affection seems to be the most popular research topics [12]. However, many of these symptoms mentioned not least the stress and emotions are heavily personal experiences. Bio-signals vary from person to person and set a collection of challenges for researcher benefiting wearables as the prior data source. Even with the migraine, which is somehow clearly demonstrable medical condition (especially if compared to stress), symptoms of patients can differ in large level. Moreover, quality and quantity of comparable cheap wrist-worn devices cannot yet reach accuracy of "real" medical instruments. However, stock devices are designed for more comfortable use in everyday basis, yet they can offer an opportunity to collect data in real-life situations unreachable by medical instruments.

In this paper, we go through our personal experiences related to challenges of data quality and quantity when the aim was to collect data using a wrist-worn wearable sensors and use it to early detection of migraine attacks [14]. However, while our example is about early detection of migraine attacks, the lessons learned can be used more widely when collecting and analyzing bio-signal data collected using wearable sensors. Some of these challenges may seem straightforward, but in our experience, that does not conclude they are always avoided or solved even by experienced researchers.

2 SENSORS TO MEASURE BIO-SIGNALS

Family of self-monitoring devices with bio-medically capable sensors include variety of devices from multiple manufacturers, included but possibly limited to smart watches, fitness trackers and bracelets, rings, pendants, and devices worn over chest. Common feature over them is skin contact, light weight, and battery dependence, usually combined with limited computing and networking capabilities making them reliable on the user's smartphone and other services to collect and analyze the data. In addition to self-monitoring functionalities, these devices often also operate as clocks, communication devices, and other smart services, which can limit their sensing capabilities and energy allowance for each sensor (i.e. causing fixed data sampling periods). Design and usability requirements can also limit capabilities and functionalities available, thus devices have to be made consumer-friendly, easy to wear, and aesthetic.

To recap, sensors integrated to these devices most commonly include accelerometer, gyroscope, magnetometer, photoplethysmography, temperature, and electrodermal activity sensors.

Photoplethysmography (PPG) sensor is used to measure blood volume pulse. Blood volume pulse can then be used to calculate heart rate and interbeat interval which is the time interval between individual beats of the heart.

Electrodermal activity (EDA) depicts skin's electrical changes caused by sweating. When sweat pores fill with sweat, the conductivity of skin increases. As the sweat secretion of skin is sympathetically controlled, sympathetic arousal can be detected as increasing skin conductance [1].

Skin temperature is measured with a infrared (IR) thermopile that detects the temperature difference between an IR absorber and a reference [10]. The skin temperature has a close connection with skin blood flow and changes in the skin temperature may affect to electrodermal activity [1].

With accelerometer, magnetometer and gyroscope the movement of the object or limb it is attached can be measured. These signals are highly dependable of the sensor location and the human in question, so their usage has to be handled carefully.

3 LESSONS LEARNED IN DATA QUALITY AND QUANTITY

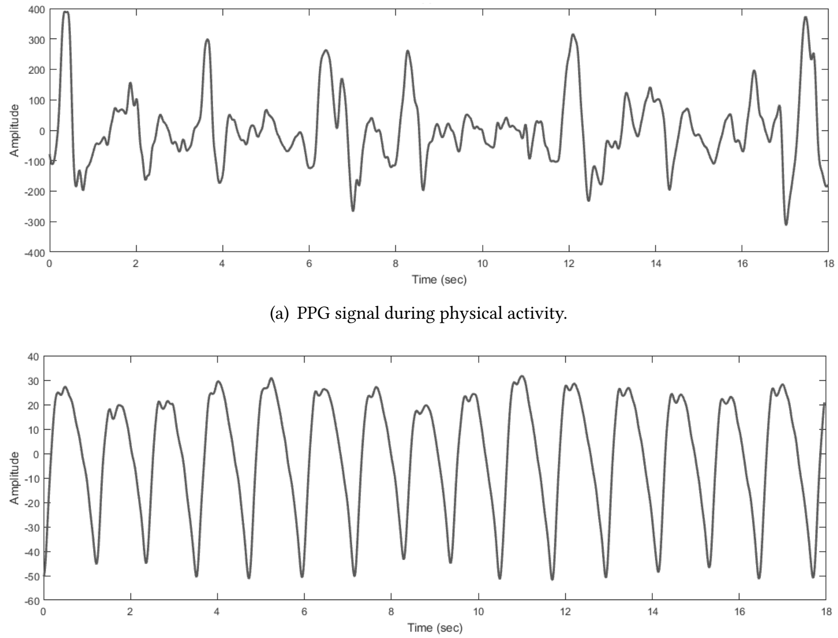
Challenges usually met by measuring patients with stock wearables can be divided into three categories: (1) usability of device, (2) amount of data, and (3) quality of data. Usability issues such as limited battery life and problems with uploading the data are often very device-specific and depend on available technologies, such as networking capabilities and protocols implemented on the device. Thus, the problems in usability are not necessarily the same through different devices.

However, manufacturers' solutions especially for battery life can have impact on the data quality and quantity, especially when sampling rate is increased or simultaneous sensor activities precluded by sake of energy consumption. Knowing these kind of issues is crucial when selecting devices for the study and comparing measured data from stock devices to actual medical devices providing a possible ground truth.

Next, we focus on more deeply in lessons learned on data quality and quantity, which may cause common pitfalls for researchers operating in the field. This paper is based on experiences we had when gathering data using a wrist-worn device to recognize migraine attacks beforehand. However, the faced challenges are not tightly related to migraine as a disease, and therefore, the same challenges should be noted when data is gathered from any other disease as well.

3.1 Quality of the Data

Without high quality training data and labels there is no high quality, reliable, recognition models. The data is of high quality when it is accurate, correct and truthful. Unfortunately the data gathered during our data collection session did not always fulfill these requirements. When predictions are made about the status of human body using wearable sensors, bio-signals such as heart rate, heart rate variability, temperature, and galvanic skin response are often in more important role than inertial sensors such as accelerometer, magnetometer, and gyroscope, which instead can be used to measure movement. However, in our previous article is was noted that physical activity can cause a lot of disturbances to bio-signals causing measurements to become unreliable or in the worst case totally missing [6].



(a) PPG signal during physical activity.

(b) PPG signal while resting.

Figure 1: Physical activity disturbes biosignal measurements.

There are at least two options how to overcome the data quality problems caused by physical activity. The first option is to leave out all the sequences of the time-series data where movement has disturbed bio-signals. For instance, in [14] the data set was preprocessed so that it includes only sleep time data. The advantage of using sleep time data is that during sleep people are mostly still, and therefore, it does not include disturbances caused by movement. This can easily be seen from Figure 1 where blood volume pulse is shown while person is moving (upper figure) and while he is sleeping (lower figure). According to Figure 1 sleep time data is much more reliable than data from active periods making it usable to detect early symptoms of migraine seizures and attacks.

Another option is to first recognize the type of human activity, and then build own bio-signal based recognition for each activity to recognize desired state from the human body. This type of approach was used for instance in [13]. In the article, the idea was to study continuous smartphone user authentication based on smartphone sensor data (including touch pressure and x- and y-coordinates of a swipe, and accelerometer data) collected from swipe-gestures. The study showed that human activity recognition improves the reliability of authentication: when a separate authentication model was trained for each human activity, authentication was more reliable than when using just a one model. While the application field of the article was not related to detecting human state based on bio-signals, it shown the potential of human activity recognition. Therefore, if there is need for detecting human state during motion and during rest, we recommend to use more than one prediction model in the recognition process.

In addition, an important aspect in data quality are the labels, without reliable labels it is not possible to train reliable models. When it comes to data that are collected 24/7 from everyday life, data cannot be labeled without information coming from study subjects. To get reliable labels, study subjects need to have clear instructions how to report about their symptoms and seizures.

In our data gathering, study subjects were instructed to keep diary about their symptoms where they described them using their own words. However, it was noted that diary entries were not unambiguous and study subjects did not always recognize early symptoms. Therefore, better idea is to make a list of possible symptoms so subjects would not need to describe symptoms, but instead, just recognize and find them from predefined list of symptoms, and rate the intensity of symptoms if needed. This approach would make labeling more uniform and unambiguous, and therefore, less vulnerable for false conclusions and errors. Moreover, to maximize the reliability of the labels, only motivated subjects should be selected for the study.

3.2 Quantity of Data

In addition to quality of data, also the amount of data is important. In order to train reliable models, a decent amount of data needs to be collected. In addition, when the aim is the early detection of seizures and attacks, data needs to be collected from healthy periods but similarly data is needed to gather during seizure and early symptoms of it. The problem from the data gathering point of view is that seizures and attacks do not occur often. For instance, in the case of migraine some patients have attacks only once a month. Therefore, the study subjects for the data gathering should be chosen so that they have attacks more frequently, for instance bimonthly, to get decent amount of data from both healthy and non-healthy days. Still this leads to very long data gathering session and highly imbalanced data set which needs to be considered in the model training phase.

Luckily, there are methods to increase the quantity of data artificially. The most obvious approach is to use re-sampling, but there are also more sophisticated methods such as SMOTE [2] and noise injection [16] that can be used to increase the amount of data. These methods do not just make copies from original samples but create new samples based on the characteristics of old one's. Therefore, they do not just increase the amount of data but also increase the variability of the dataset. Another approach was presented in [14] where in the first place feature from sleep data were extracted. Then secondly, data size was increased by comparing features from nights before a non-seizure day pairwise, and comparing nights before a seizure day to nights before a non-seizure day. This procedure was shown to increase the number of observations remarkable.

Another factor that needs to be noted is that not all seizures and attacks are the same, in fact, symptoms can be highly personal. For instance, migraine types can be divided into six main categories, and each of these into several sub-categories. Therefore, in order to train user-independent recognition model, data needs to be collected from a huge number of study subjects. Second option is to train personal recognition model for each subject. This approach requires data gathering session from each new user which means that device the detect seizures and attacks beforehand would not be usable out-of-the-box. Third option is to personalize models based on online data [15], this can be done for instance based on methods of incremental learning [5]. The idea of this type of model is that in the first phase recognition is based on user-independent model but when online data is available, model can be updated, and therefore, personalized based on personal data to better predict subject's individual features. The advantage of incremental learning compared to many other methods is that it does not require model re-training, instead models can be updated by demand.

Moreover, when it comes to the amount of the data, the whole research community should work together by releasing the collected data sets publicly available. This would not only help to solve the problem of too small data sets, but it would also help to validate and compare the used methods. In fact, some data sets already are publicly available. For instance, the data used in [11] which contains bio-signal data collected using Empatica E4-device [3] for stress and affect state recognition from 15 persons.

4 DISCUSSION AND CONCLUSION

Early diagnosis of seizures and attacks using wearable sensors has a huge market potential, especially if and when recognition of conditions can by done by cheap stock devices. However, these same cheap stock wearables and their use in real-life conditions can cause variety of challenges for researchers collecting the data and building diagnosis models. In this paper, we have especially discussed our lessons learned with building migraine prediction method by using wrist-worn wearables.

To build reliable recognition models, we need an extensive and high quality training data set that can be expensive and hard to collect. This study concentrated on challenges of data quality and quantity especially caused by using multifunctional stock wearables instead of expensive medical instruments. For instance, movement of the patient causes a lot of disturbances to bio-signals making them unreliable. Therefore, we suggested to base the recognition to sleep time data as people are mainly still while sleeping. If the data are labeled by study subjects themselves on daily basis, to obtain uniform labels it is better idea to ask subjects to find symptoms from predefined set than to ask them to describe them using own words. However, by our experience, wearables can provide us real-life data from patients' everyday life impossible to collect in laboratory conditions.

Moreover, it was noted that the amount of data can be an issue when analyzing rare or non-frequent symptoms and conditions. Seizures and attacks do not occur very often making data set imbalanced and data gathering sessions long. However, there are solutions to increase the amount of data artificially. In addition, symptoms can be highly personal and vary through patients a lot. Therefore, data should be collected from a huge number of study subjects if the aim is to build user-independent recognition model. Other option is to build personal or personalized recognition model that may, indeed, with some symptoms be more accurate approach than aiming for a general model.

To conclude, we encourage wearable community to share their lessons learned with data collection, analytics, and research studies to provide a full picture of challenges in our field. Trial and error -based approaches can lead on expensive failures, thus recognizing challenges and providing solutions becomes essential for future work. It is clear that stock wearables can provide extensive data from real-life conditions, but challenges seen especially in processing such a data has to be noted and potential risks for research methods discussed through the community.

ACKNOWLEDGEMENT

This research is supported by the Academy of Finland 6Genesis Flagship (grant 318927) and Business Finland funding for Reboot IoT Factory-project[1]. Authors are also thankful for Infotech Oulu and Nokia Foundation (Jorma Ollila Grant for Dr Ella Peltonen).

REFERENCES

[1] W. Boucsein. 2012. *Electrodermal Activity*. Springer, New York.
[2] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority oversampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.
[3] Empatica E4. 2016. https://www.empatica.com/e4-wristband. Accessed: 2016-06-10.
[4] K. Fujiwara, M. Miyajima, T. Yamakawa, E. Abe, Y. Suzuki, Y. Sawada, M. Kano, T. Maehara, K. Ohta, T. Sasai-Sakuma, T. Sasano, M. Matsuura, and E. Matsushima. 2016. Epileptic Seizure Prediction Based on Multivariate Statistical Process Control of Heart Rate Variability Features. *IEEE TBME* 63, 6 (June 2016), 1321–1332.
[5] A. Gepperth and B. Hammer. 2016. Incremental learning algorithms and applications. In *ESANN 2016*.
[6] Heli Koskimäki, Henna Mönttinen, Pekka Siirtola, Hanna-Leena Huttunen, Raija Halonen, and Juha Röning. 2017. Early Detection of Migraine Attacks Based on Wearable Sensors: Experiences of Data Collection Using Empatica E4 *(UbiComp '17)*. ACM, 506–511. https://doi.org/10.1145/3123024.3124434

[1] www.rebootiotfactory.fi

[7] Chikao Nakayama, Koichi Fujiwara, Masahiro Matsuo, Manabu Kano, and Hiroshi Kadotani. 2015. Development of sleep apnea syndrome screening algorithm by using heart rate variability analysis and support vector machine. In *EMBC 2015*. IEEE, 8165–8168.

[8] Q. W. Oung, M. Hariharan, H. L. Lee, S. N. Basah, M. Sarillee, and C. H. Lee. 2015. Wearable multimodal sensors for evaluation of patients with Parkinson disease. In *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*. 269–274.

[9] J. Pagan, D. Orbe, M. Irene, A. Gago, M. Sobrado, J. L. Risco-Mart?n, and J. L. Ayala. 2015. Robust and Accurate Modeling Approaches for Migraine Per-Patient Prediction from Ambulatory Data. *Sensors* 15, 7 (2015), 15419–15442.

[10] Vasco Polyzoev and Habib Karaki. 2016. Sensors Mag Demystifying Thermopile IR Temp Sensors. http://www.sensorsmag.com/sensors-mag/demystifying-thermopile-ir-temp-sensors-13157. Accessed: 2016-06-10.

[11] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 400–408.

[12] Philip Schmidt, Attila Reiss, Robert Duerichen, and Kristof Van Laerhoven. 2018. Wearable affect and stress recognition: A review. *arXiv preprint arXiv:1811.08854* (2018).

[13] Pekka Siirtola, Jukka Komulainen, and Vili Kellokumpu. 2018. Effect of context in swipe gesture-based continuous authentication on smartphones. In *26th European Symposium on Artificial Neural Networks, ESANN 2018, Bruges, Belgium, April 25-27, 2018*. 639–644.

[14] Pekka Siirtola, Heli Koskimäki, Henna Mönttinen, and Juha Röning. 2018. Using Sleep Time Data from Wearable Sensors for Early Detection of Migraine Attacks. *Sensors* 18, 5 (2018). https://doi.org/10.3390/s18051374

[15] P. Siirtola, H. Koskimäki, Röning, and J. 2018. Personalizing human activity recognition models using incremental learning. In *26th European Symposium on Artificial Neural Networks, ESANN 2018, Bruges, Belgium, April 25-27, 2018*. 627—-632.

[16] Pekka Siirtola, Heli Koskimäki, and Juha Röning. 2016. Personal models for eHealth-improving user-dependent human activity recognition models using noise injection. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*. IEEE, 1–7.

[17] LJ Stovner, J-A Zwart, K Hagen, GM Terwindt, and J. Pascual. 2006. Epidemiology of headache in Europe. *European Journal of Neurology* 13 (2006), 333?345.