Improving Heart Rate Variability Measurements from Consumer Smartwatches with Machine Learning

Martin Maritsch
mmaritsch@ethz.ch
ETH Zurich
Zurich, Switzerland

Caterina Bérubé
cberube@ethz.ch
ETH Zurich
Zurich, Switzerland

Mathias Kraus
mathiaskraus@ethz.ch
ETH Zurich
Zurich, Switzerland

Vera Lehmann
vera.lehmann@insel.ch
Inselspital, Bern University Hospital
Bern, Switzerland

Thomas Züger
thomas.zueger@insel.ch
Inselspital, Bern University Hospital
Bern, Switzerland

Stefan Feuerriegel
sfeuerriegel@ethz.ch
ETH Zurich
Zurich, Switzerland

Tobias Kowatsch[*]
tobias.kowatsch@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

Felix Wortmann
felix.wortmann@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

ABSTRACT

The reactions of the human body to physical exercise, psychophysiological stress and heart diseases are reflected in heart rate variability (HRV). Thus, continuous monitoring of HRV can contribute to determining and predicting issues in well-being and mental health. HRV can be measured in everyday life by consumer wearable devices such as smartwatches which are easily accessible and affordable. However, they are arguably accurate due to the stability of the sensor. We hypothesize a systematic error which is related to the wearer movement. Our evidence builds upon explanatory and predictive modeling: we find a statistically significant correlation between error in HRV measurements and the wearer movement. We show that this error can be minimized by bringing into context additional available sensor information, such as accelerometer data. This work demonstrates our research-in-progress on how neural learning can minimize the error of such smartwatch HRV measurements.

[*]Also with ETH Zurich.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous computing**; • **Applied computing** → **Consumer health**; • **Computing methodologies** → *Neural networks.*

KEYWORDS

neural networks, heart rate variability, smartwatch

1 INTRODUCTION

The ability of human heart to adapt to sudden physiological changes is visible in heart rate variability (HRV); the variation of subsequent inter-beat intervals across time. In fact, it informs about involuntary physiological functions, as it is a valid measure of the interaction between the autonomic nervous systems (i.e. sympathetic and parasympathetic). HRV reflects individual reactions to physical exercise, psychophysiological stress and heart diseases. In particular, HRV has been associated with parasympathetic activity [4, 11] and hypoglycemic episodes [2, 22] of type 1 diabetes mellitus patients, which are caused by a deficit in the sympathetic nervous system [3].

Self-tracking devices are becoming increasingly popular: sales of sports and fitness trackers increased from 97.6 million in 2015 to 134 million in 2018 and are expected to reach 148.5 million in 2021 [13] (cf. also [12] for a review on availability of wrist-worn fitness wearable devices and sensors). Due to their discrete form-factor, smartwatches are suitable for continuous monitoring of the wearers vital factors. To this end, they are often equipped with measurement capabilities such as an optical heart rate sensor. Besides observing the heart rate, raw data from these sensors can as well be used to compute HRV. Their affordability and unobtrusiveness makes smartwatches relevant to a large audience and, in particular, more likely to be adapted than professional electrocardiographies (ECGs) for measuring HRV.

Wearables measuring HRV can considerably contribute to improving well-being and mental health. For instance, it can predict acute complications of type 1 diabetes mellitus (i.e., hypoglycemia). As such, its benefit to better health becomes relevant to around 40 million people worldwide and this number is expected to double within the next 10 to 20 years [8].

Objective: This work demonstrates research-in-progress on the potential of machine learning approaches improving smartwatch HRV measurements. More specifically, we aim to utilize explanatory and predictive modeling to the sensor data provided by smartwatches with the goal of minimizing errors in HRV measurements due to physical activity.

2 RELATED WORK

Consumer Wearable Devices

Wearable devices allow non-invasive monitoring of physiological activity. In particular, smartwatches can be applied to different contexts, such as stress [4], sleep quality [19] or physical fitness [5] to measure well-being through heart rate. Moreover, these devices are often equipped with a range of additional sensors, allowing for more holistic measurements. For instance, there are wrist-based consumer smartwatches that provide multiple data dimensions such as inter-beat intervals (obtained via optical sensor), three-axis accelerometer data, steps, burned calories or proprietary stress values. However, such non-professional devices need to be compared to a more precise instrument. The *Firstbeat Bodyguard 2 (heart rate monitor)* is an inter-beat interval recorder employing two electrodes on the chest for measurement and that can be considered as a semi-professional device [18].

Smartwatches thus collect a magnitude of data, most of which is sufficient for giving their user an overview of their daily activity. However, when being applied to more serious medical and healthcare use cases, the current measurement accuracy of wearable devices remains rather unclear [5, 6, 9, 10, 20].

Heart Rate Variability

HRV measurements are relevant in various applications. For example, the study conducted in [2] used an ECG to derive HRV for hypoglycemia prediction and detection. However, to reach a broader audience, we propose to use consumer wearable devices such as smartwatches instead of professional ECGs to measure HRV. While some studies show that data from smartwatches was highly accurate when compared with professional ECGs in long-term measurements [19], other previous research observed a discrepancy between measurements from wrist-worn trackers and ECGs [5, 10].

Resolution: HRV can be measured in both the time and frequency domain. According to the Task Force of the European Society of Cardiology [15], short-term recordings (i.e. 2 to 5 minutes) should be assessed with standard deviation of normal-to-normal interval (SDNN) and root mean square of the successive differences (RMSSD) in the time domain. However, RMSSD has been found to be reliable when calculated with a sample of 10, 30 or 60 seconds, which was not always the case for SDNN, or frequency-based measures [1, 7, 16, 17, 21, 23]. Ultra-short-term measures of HRV such as RMSSD would thus allow monitoring streams of physiological changes with a relatively high resolution.

HRV and hypoglycemia: HRV has been shown to correlate positively with hypoglycemia [22], although in a short-term measure (i.e., 5 min), and with continuous glucose monitoring as a valid predictor of hypoglycemia [2]. It is therefore in our interest to explore the possibility of using an ultra-short-term measurement of HRV to predict hypoglycemia.

3 WORK-IN-PROGRESS RESEARCH

Problem Statement

Many widespread smartwatches are capable of measuring HRV, however, our analysis has shown that they incur systematic measurement errors leading to inaccurate HRV measurements. We hypothesize that these data quality issues are related to the movement or physical activity of the wearer and the device being unable to properly measure inter-beat intervals during these times.

Smartwatches are nowadays equipped with a range of sensors relevant for health measurement, such as for example, an optical heart sensor, an accelerometer, a compass, or sensors for positioning information. These sensors can potentially provide valuable additional information.

Our aim is to minimize the error of smartwatch HRV measurements by using additional available sensor information for correcting HRV measurements. With our work, we want to improve the quality and reliability of the RMSSD calculated by smartwatches. The basis of these calculations are the inter-beat intervals as measured by those devices.
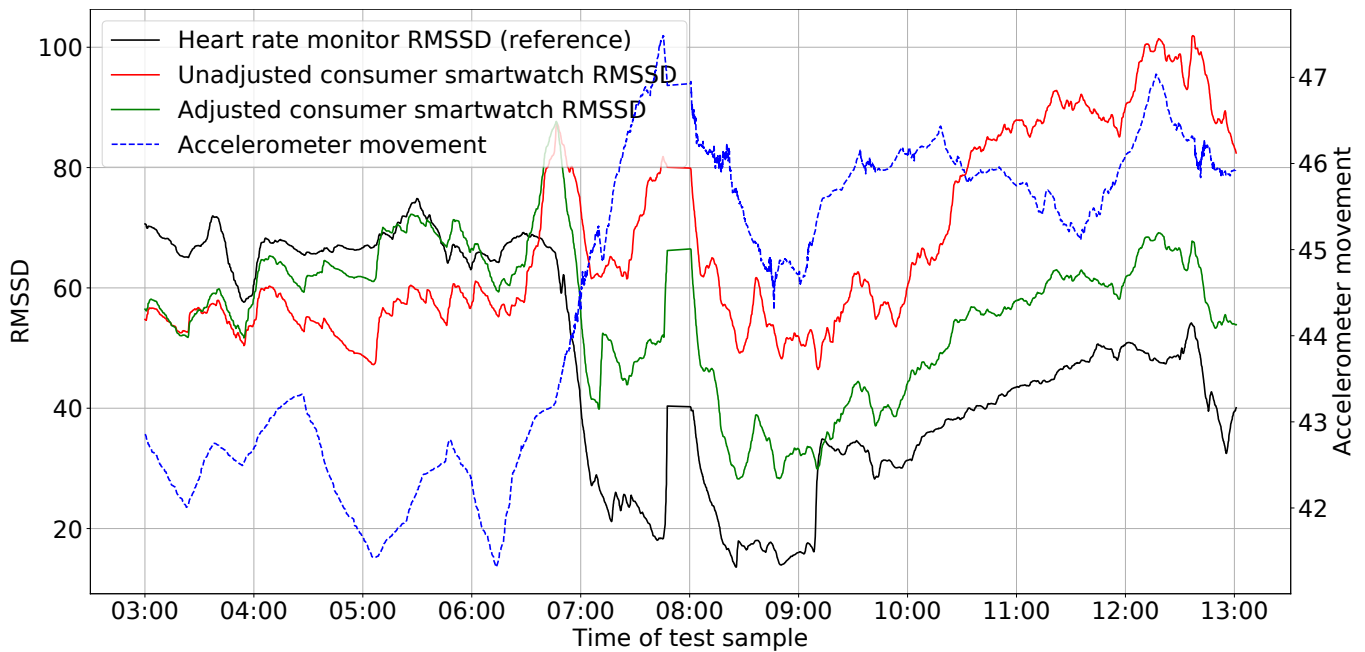
Figure 1: Samples of the RMSSD as calculated from data of the heart rate monitor (black, reference value) and the consumer smartwatch (consumer smartwatch) (red, unadjusted). The line in green shows the adjusted consumer smartwatch RMSSD, which is the sum of the unadjusted RMSSD and the error predicted by the neural network. The dashed blue line indicates accelerometer movement. Values have been smoothed to facilitate visualization.

Methods

Devices: In our preliminary research, we compare data obtained from a chest-based heart rate monitor (*Firstbeat Bodyguard 2*) with data from one of the most widespread consumer smartwatches which is equipped with an accelerometer and an optical heart rate sensor capable of measuring inter-beat intervals.

The heart rate monitor allows recording inter-beat intervals by means of a two-lead measurement, thus providing reference HRV data for our investigations. Given the cited related work on smartwatches we consider obtained from the consumer smartwatch to be less accurate. However, the consumer smartwatch provides further data dimensions, which we take into account in order to compensate for the inaccuracies in HRV measurements imposed by inaccurate inter-beat interval data.

For instance, we aim to improve the accuracy of RMSSD calculations by including both inter-beat intervals and movement indices such as orientation and angular velocity data from accelerometer sensors into the corresponding calculations.

Analysis: In a first step, we use explanatory analysis to compare and investigate the differences between the HRV measurements from both devices, and how the differences

relate to the wearer movement. Then, in a second step, we apply predictive modeling in order to forecast measurement errors imposed by the described wearer movements.

Prediction Model: In order to estimate the aforementioned measurement error, we make use of a convolutional neural network (CNN). CNNs are a sub-group of neural networks, which are commonly used for classification tasks like object recognition in images [14]. In particular, CNNs have also successfully been applied to medical time-series data such as for ECG classification [24].

In our model, we frame the obtained data as a time-series task in a supervised learning setting. The aim is to estimate the errors between the consumer smartwatch RMSSDs (*samples* in the machine learning model) and the reference heart rate monitor RMSSDs (*labels* in the machine learning model).

Data

Raw data was obtained from the heart rate monitor via the *Firstbeat SPORTS Individual* software. From the consumer smartwatch, data was gathered via a custom smartphone app utilizing the Bluetooth streaming capability of the smartwatch.

Data was recorded by a healthy individual wearing both devices simultaneously over a period of 72 hours. After preprocessing, a total of roughly 200,000 observations consisting

of heart rate monitor inter-beat interval, consumer smart-watch inter-beat interval and consumer smartwatch three-axis accelerometer data were left. We split the dataset into 80% of the observations for training and 20% for testing.

4 RESULTS

We quantify HRV by RMSD which is calculated as

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(IBI_i - IBI_{i-1})^2},$$

where N is the number of samples within a sliding window of 60 seconds and IBI_i is the i-th inter-beat interval within those samples.

The error of the measurements between the heart rate monitor and the consumer smartwatch at time t is defined as

$$\varepsilon_t = RMSSD_{\text{heart rate monitor}_t} - RMSSD_{\text{consumer smartwatch}_t}.$$

Explanatory Analysis

We first tested for a tendency to systematic errors in HRV measurements of the consumer smartwatch, which are imposed by external influences such as the wearer movement. As a result, we found a statistically significant positive correlation between accelerator movement and measurement error, $r = 0.22$, $p = \leq 0.001$.

Predictive Analysis

Error Prediction Model: In our approach, we used a deep CNN in order to predict the error of the measurement. Subsequently, the predicted error is added to the RMSSD as calculated from the raw inter-beat intervals from the consumer smartwatch in order to get an approximation of the reference RMSSD calculated from the heart rate monitor data.

Performance: The first experiment yielded an improvement of the root-mean-squared error (RMSE) between heart rate monitor and consumer smartwatch RMSSDs from an initial 48.89 down to 28.50 on the test set. The statistically significant correlation of heart rate monitor and consumer smartwatch RMSSDs in the test set was improved from $r = 0.37$, $p = \leq 0.001$ before adjusting to $r = 0.58$, $p = \leq 0.001$ after adjusting.

Figure 1 visualizes this reduction of the measurement error in RMSSDs. For the first part of the graph, the RMSSD error is rather low. During this time, the wearer was in bed, which is reflected by low accelerometer movement. Later on, at around 06:30 when the wearer got out of bed, a rise of accelerometer movement and a deterioration of the error in raw RMSSD measurements are observable. Additionally, it can be observed how the addition of the error as predicted by the CNN turns from positive into negative. The sensor data

suggests that a larger magnitude of accelerometer movement is related to a larger error in measurement.

5 CONCLUSION

Prediction of HRV measurements with smartwatches would allow non-invasive continuous monitoring of psychophysiological conditions such as the risk of hypoglycemia. The key benefit of being able to accurately measure HRV with smartwatches would certainly be their popularity and easiness of use in contrast to professional devices. The implemented CNN is capable of reducing the error in consumer smartwatch HRV measurements by taking into account additional motion information in the form of accelerometer data.

Current Limitations: In future research, we will investigate the generalizability of our approach, which needs to be done in a larger-scale study. Furthermore, the restricted accuracy of the heart rate monitor (see [18]) also limits the accuracy of the model evolved on top of its data. We thus aim to conduct a larger-scale study in which we gather reference data with a professional-grade ECG device.

Outlook: We found strong evidence that systematic errors in HRV measurements from the consumer smartwatch can be minimized with the utilization of additional data by the use of neural networks. In the future, we can generate additional data dimensions that potentially carry valuable information for our problem (e.g., the time-shifted difference of accelerometer data) by utilizing methods of feature engineering. While we are aware of that our proposed methods will not be capable of completely eliminating the error in smartwatch HRV measurements, the methods we have shown contribute to more reliable measurements of physiological values such as HRV in smartwatches. Furthermore, even more, widespread smartwatches such as the *Apple Watch* should be investigated.

REFERENCES

[1] Hyun Jae Baek, Chul-Ho Cho, Jaegeol Cho, and Jong-Min Woo. 2015. Reliability of Ultra-Short-Term Analysis as a Surrogate of Standard 5-Min Analysis of Heart Rate Variability. *Telemedicine and e-Health* 21, 5 (2015), 404–414. https://doi.org/10.1089/tmj.2014.0104

[2] Simon Lebech Cichosz, Jan Frystyk, Ole K. Hejlesen, Lise Tarnow, and Jesper Fleischer. 2014. A novel algorithm for prediction and detection of hypoglycemia based on continuous glucose monitoring and heart rate variability in patients with type 1 diabetes. *Journal of Diabetes Science and Technology* 8, 4 (2014), 731–737. https://doi.org/10.1177/1932296814528838

[3] Philip E. Cryer. 2004. Diverse Causes of Hypoglycemia-Associated Autonomic Failure in Diabetes. *New England Journal of Medicine* 350, 22 (2004), 2272–2279. https://doi.org/10.1056/NEJMra031354

[4] D.A. Dimitriev and E.V. Saperova. 2015. Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing mental stress. *Autonomic Neuroscience* 192, 3 (2015), 80. https://doi.org/10.1016/j.autneu.2015.07.086

[5] Erin E Dooley, Natalie M Golaszewski, and John B Bartholomew. 2017. Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. *JMIR mHealth and uHealth* 5, 3 (2017), e34. https://doi.org/10.2196/mhealth.7043

[6] Fatema El-Amrawy and Mohamed Ismail Nounou. 2015. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthcare Informatics Research* 21, 4 (2015), 315–320. https://doi.org/10.4258/hir.2015.21.4.315

[7] Michael R Esco and Andrew A. Flatt. 2014. Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: Evaluating the agreement with accepted recommendations. *Journal of Sports Science and Medicine* 13, 3 (2014), 535–541.

[8] International Diabetes Federation. 2017. *IDF Diabetes Atlas* (8 ed.). International Diabetes Federation. http://diabetesatlas.org

[9] Z. Ge, P. W.C. Prasad, N. Costadopoulos, Abeer Alsadoon, A. K. Singh, and A. Elchouemi. 2016. Evaluating the accuracy of wearable heart rate monitors. In *Proceedings - 2016 International Conference on Advances in Computing, Communication and Automation (Fall), ICACCA 2016*. IEEE, 1–6. https://doi.org/10.1109/ICACCAF.2016.7748986

[10] A. Marc Gillinov, Muhammad Etiwy, Stephen Gillinov, Robert Wang, Gordon Blackburn, Dermot Phelan, Penny Houghtaling, Hoda Javadikasgari, and Milind Y. Desai. 2017. Variable Accuracy of Commercially Available Wearable Heart Rate Monitors. *Journal of the American College of Cardiology* 69, 11 (2017), 336. https://doi.org/10.1016/s0735-1097(17)33725-7

[11] Jeffrey J. Goldberger, Sridevi Challapalli, Roderick Tung, Michele A. Parker, and Alan H. Kadish. 2001. Relationship of heart rate variability to parasympathetic effect. *Circulation* 103, 15 (2001), 1977–1983. https://doi.org/10.1161/01.CIR.103.15.1977

[12] André Henriksen, Martin Haugen Mikalsen, Ashenafi Zebene Woldaregay, Miroslav Muzny, Gunnar Hartvigsen, Laila Arnesdatter Hopstock, and Sameline Grimsgaard. 2018. Using fitness trackers and smartwatches to measure physical activity in research: Analysis of consumer wrist-worn wearables. *Journal of Medical Internet Research* 20, 3 (2018), e110. https://doi.org/10.2196/jmir.9157

[13] Mordor Intelligence. 2018. Smart Watch Market - Growth, Trends, and Forecast (2019 - 2024). https://www.mordorintelligence.com/industry-reports/global-smart-watches-market-industry

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in neural information processing systems*. 1097–1105.

[15] Marek Malik. 1996. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *European Heart Journal* 17, 3 (1996), 354–81. http://www.ncbi.nlm.nih.gov/pubmed/8737210

[16] M. Loretto Munoz, Arie Van Roon, Harriëtte Riese, Chris Thio, Emma Oostenbroek, Iris Westrik, Eco J.C. De Geus, Ron Gansevoort, Joop Lefrandt, Ilja M. Nolte, and Harold Snieder. 2015. Validity of (Ultra-)Short recordings for heart rate variability measurements. *PLoS ONE* 10, 9 (2015), e0138921. https://doi.org/10.1371/journal.pone.0138921

[17] Udi Nussinovitch, Keren Politi Elishkevitz, Keren Katz, Moshe Nussinovitch, Shlomo Segev, Benjamin Volovitz, and Naomi Nussinovitch. 2011. Reliability of ultra-short ECG indices for heart rate variability. *Annals of Noninvasive Electrocardiology* 16, 2 (2011), 117–122. https://doi.org/10.1111/j.1542-474X.2011.00417.x

[18] Jakub Parak and Ilkka Korhonen. 2013. Accuracy of Firstbeat Bodyguard 2 beat-to-beat heart rate monitor. *(Whitepaper)* (2013), 6–8. https://assets.firstbeat.com/firstbeat/uploads/2015/10/white_paper_bodyguard2_final.pdf

[19] Dung Phan, Lee Yee Siong, Pubudu N. Pathirana, and Aruna Seneviratne. 2015. Smartwatch: Performance evaluation for long-term heart rate monitoring. In *4th International Symposium on Bioelectronics and Bioinformatics, ISBB 2015*. 144–147. https://doi.org/10.1109/ISBB.2015.7344944

[20] Dung Phan, Lee Yee Siong, Pubudu N Pathirana, and Aruna Seneviratne. 2015. Smartwatch: Performance evaluation for long-term heart rate monitoring - IEEE Conference Publication. In *Ieeexplore.Ieee.Org*. IEEE, 144–147. https://ieeexplore.ieee.org/document/7344944/

[21] Lizawati Salahuddin, Jaegeol Cho, Myeong Gi Jeong, and Desok Kim. 2007. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*. IEEE, 4656–4659. https://doi.org/10.1109/IEMBS.2007.4353378

[22] Hartmut Schächinger, Johannes Port, Stuart Brody, Lilly Linder, Frank H. Wilhelm, Peter R. Huber, Daniel Cox, and Ulrich Keller. 2004. Increased high-frequency heart rate variability during insulin-induced hypoglycaemia in healthy humans. *Clinical Science* 106, 6 (2004), 583–588. https://doi.org/10.1042/cs20030337

[23] T. Thong, K. Li, J. McNames, M. Aboy, and B. Goldstein. 2004. Accuracy of ultra-short heart rate variability measures. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, Vol. 3. IEEE, 2424–2427. https://doi.org/10.1109/iembs.2003.1280405

[24] Martin Zihlmann, Dmytro Perekrestenko, and Michael Tschannen. 2017. Convolutional Recurrent Neural Networks for Electrocardiogram Classification. (2017), 1–4. arXiv:1710.06122 http://arxiv.org/abs/1710.06122