
Learning Individualized Cardiovascular Responses from Large-scale Wearable Sensors Data

Haraldur T. Hallgrímsson*
Department of Computer Science
University of California, Santa Barbara
hth@cs.ucsb.edu

Filip Jankovic
Evidation Health
Santa Barbara, CA
fjankovic@evidation.com

Tim Althoff
Paul G. Allen School of Computer Science & Engineering
University of Washington
althoff@cs.washington.edu

Luca Foschini
Evidation Health
Santa Barbara, CA
luca@evidation.com

Abstract

We consider the problem of modeling cardiovascular responses to physical activity and sleep changes captured by wearable sensors in free living conditions. We use an attentional convolutional neural network to learn parsimonious signatures of individual cardiovascular response from data recorded at the minute level resolution over several months on a cohort of 80k people. We demonstrate internal validity by showing that signatures generated on an individual’s 2017 data generalize to predict minute-level heart rate from physical activity and sleep for the same individual in 2018, outperforming several time-series forecasting baselines. We also show external validity demonstrating that signatures outperform plain resting heart rate (RHR) in predicting variables associated with cardiovascular functions, such as age and Body Mass Index (BMI). We believe that the computed cardiovascular signatures have utility in monitoring cardiovascular health over time, including detecting abnormalities and quantifying recovery from acute events.

1 Introduction

When engaging in any physical task, the human body responds through a series of integrated changes in function that involves its physiologic systems, including the musculoskeletal, the cardiovascular, and the respiratory systems [5]. Such responses may vary significantly due to environmental factors, yet when elicited in a controlled environment such as a 6-minute walk test carried out in lab settings, they allow inferring individual-specific physiological markers such as Resting Heart Rate (RHR), Maximal Heart Rate, and Maximal Stroke Value. These markers are important in characterizing an individual’s health and fitness status. For example, it is well known that cardio-respiratory fitness is inversely associated with all-cause mortality [8].

Recently, the advent and widespread adoption of wearable devices and fitness tracking apps [13] has enabled continuous, unobtrusive tracking of an individual behavior and physiological signals such as heart rate, physical activity, and sleep over time, with time resolution down to the minute-level and below. This has enabled population-scale physiological sensing [2].

In this work, we move beyond population-level aggregated sensing and set out to learn *individual* characteristics of cardiovascular responses by observing the relationship between behaviors such as

*Work done while interning at Evidation Health.

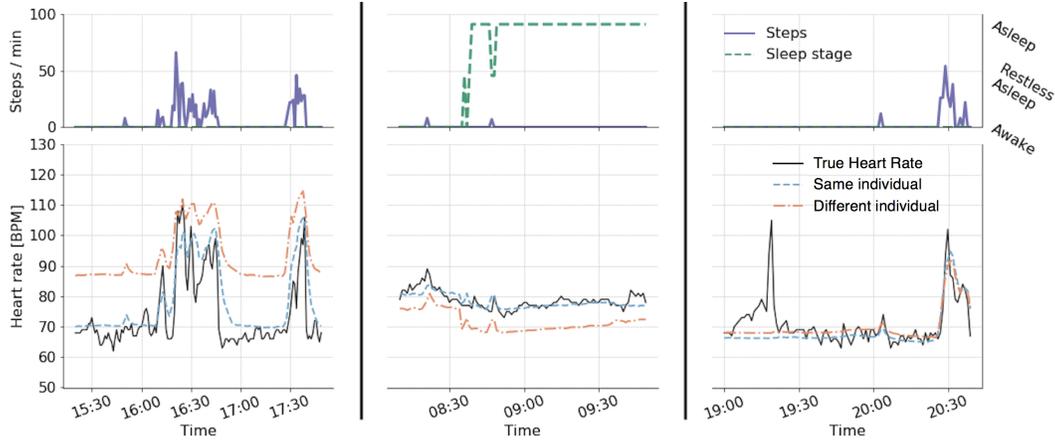


Figure 1: Example physical activity and sleep (upper row) and heart rate (bottom row) sensor data from three individuals, demonstrating how heart rate responds to onset of exercise (left column) and sleep (middle column). Changes in heart rate do not always occur due to physical activity (right column), with onset of anxiety or stress being potential unmeasured confounders. As expected, applying a signature from a different person (demonstrated in orange) results in increased reconstruction error.

sleep and physical activity and their associated changes in heart rate during the individuals everyday life. In absence of the controlled lab settings usually described in the physiology literature [5], we hypothesize that prolonged observation periods increase the likelihood of a behavior mimicking an in-lab test to spontaneously occur. For example, a brisk walk to the train station may be a good approximation of a 6-minute walk test. For this reason, we make use of attentioned models to pick up on such “natural experiments” that collectively contribute to shaping the envelope of an individual’s cardiovascular response. In an analogy with control theory, we set out to learn the cardiovascular *transfer function* of an individual to capture the cardiac output for each possible (behavioral) input.

Though previous studies have leveraged representation learning to extract health-related features from wearable sensor data [14, 3], our work is unique in terms of dataset size (2.6×10^9 minutes of sensor measurements considered from 80k users over a span of one year), outputs (parsimonious individualized cardiovascular signatures output by attentioned convolutional autoencoders), and validation methods. We believe that accurately capturing cardiovascular response enables screening for abnormalities and detecting physiological changes over time unobtrusively in free living conditions.

2 Data

We select a cohort of 80,137 members of Achievement, a commercial reward platform. To be included in this study, users must have authorized sharing with Achievement of dense minute-level steps/sleep/heart rate activity logs from commercial activity trackers, such as Fitbit or Apple Watch. Following [11], to be included in the cohort a member must have at least 10 days worth of physical activity logs, with no more than 4 hours of unreported data per day, for one or both of the collection windows of January 2017 or 2018. Half of the members reported between 26,488 to 40,537 minutes per month, averaging 32,750 minutes. 82.8% of this cohort is female, with a median age and BMI of 31 and 28.3, respectively. All members with reported data in both of the two months were assigned to the validation set (N=25,406). The remaining individuals were randomly assigned to either the training (N=43,784) or tuning sets (N=10,947).

The data from the activity trackers are minute-level measurements of a person’s total step count and average heart rate, and if the wearer is asleep or restless asleep; see Figure 1 for sample data from three individuals. We scaled these measurements to the range (0, 1) to speed up model training [10]: the heart rate per-person is whitened to zero mean and unit standard deviation, and the step count values are log-transformed to handle the large spread of values as: $\text{steps}' = \log(\text{steps} + 1) / 5$. The two sleep stages are encoded as separate binary channels. Missing data is imputed as mean heart rate of activity at awake, and no other data cleaning is performed.

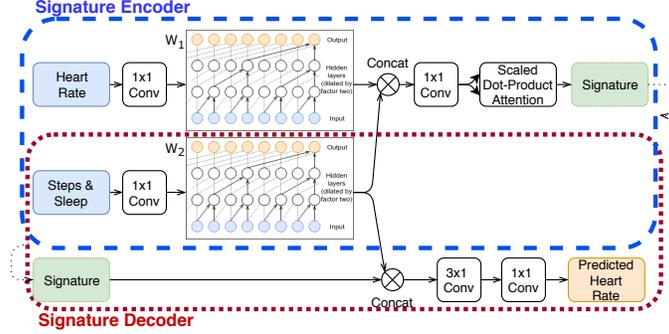


Figure 2: Diagram of proposed model architecture. The signature encoder predicts a cardiovascular signature from measured sensor data (top dashed box), and the signature decoder uses that same signature as well as physical activity data to predict the heart rate (bottom dotted box).

3 Cardiovascular Signature Network

To learn a personalized cardiovascular response function, we consider a heart rate autoencoder [4] that is conditioned on the physical activity and sleep stages. The signature-encoder learns a signature of a person based on how their heart rate responds to physical activity, while the signature-decoder uses a learned signature to predict a person’s heart rate based on their physical activity.

Encoder: The encoder model, as seen in Figure 2, learns a fixed-size signature from an arbitrarily length time-series. It consists of two WaveNet [15] convolutional neural network (CNN) blocks, W_1 and W_2 , composed of seven dilated causal convolutional layers with residual connections and allow for modeling long-range temporal dependencies of up to 128 minutes, with 32 and 16 filters per layer, respectively. As opposed to recurrent layers, convolutions are typically faster to train especially when applied to very large sequences such as considered here. The encoder considers the physical activity channels and the heart rate signal separately in W_1 and W_2 , which allows the encoder to jointly learn a latent physical activity representation with the decoder by sharing the weights of W_2 . The outputs of W_1 and W_2 are concatenated together and a scaled dot-product attention mechanism [16] is applied to predict the cardiovascular signature with queries and keys of dimension $d_k = d_v = 8$ while the dimensionality of the values, d_v , is equal that of the signature size. Three separate convolutional layers of filter width 1 are applied to re-size the tensors appropriately.

Decoder: The decoder model consists of a single WaveNet block W_2 , whose weights are tied to that of the encoder’s, followed by two temporal convolutional layers. The output of W_2 at every time step is concatenated with a signature vector, and two temporal convolutional layers are then trained to predict the corresponding heart rate signal. The number of parameters unique to the decoder are kept to a minimum to force the signature to be as informative as possible.

Training: The two models are learned end to end by minimizing the average L_2 norm of the error in predicting heart rate, using the Adam optimizer [9] with default parameters ($\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-7}$). Missing data is imputed as mean heart rate of no activity at awake, though no loss is propagated corresponding to these periods. The models are implemented in Keras [7] with a TensorFlow [1] backend. All hidden layers include ReLU activation functions [12], with the exception of the WaveNet blocks, which use gated activation units [15], and the output, which has no non-linearity. Training was done on mini-batches of size 16, for up to 30 epochs with an early stopping criteria if validation error was not observed to improve for five epochs.

4 Experimental results

Baseline models: We consider three baselines to compare our model to. The simplest predicts a persons mean heart rate at awake or asleep. The second uses XGBoost [6] with default parameters, trained on a single person to predict their heart rate based on the previous 120 minutes of physical activity. The third uses XGBoost again, but this time trained on a population of people rather than at the individual level. The performance of the baseline models can be seen in Table 1. Both XGBoost models are trained on the January 2017 activity window and validated on January 2018.

Table 1: Experimental results. The trained proposed model was validated on the 2017 data, and also using 2017 signatures applied to 2018 data. While varying signature sizes (results shown in middle column) the full training set was used, and when varying training set size (results shown in right column) a size-32 signature was used. For comparison, we include the performance of the baseline models (in left column) trained on 2017 data and validated on 2018 data.

Baseline model	Validation error	Varying signature size			Varying training set size		
		Size	Validation error 2017 / 2018		# people	Validation error 2017 / 2018	
Awake/Sleep Mean	0.755	4	0.295	0.385	500	0.319	0.400
Individual XGBoost	0.445	8	0.291	0.385	2,000	0.306	0.391
Population XGBoost	0.539	16	0.283	0.394	5,000	0.298	0.383
		32	0.279	0.393	20,000	0.285	0.387
		64	0.288	0.384	43,784	0.279	0.393
		128	0.278	0.395			

Sensitivity analysis on signature size: We trained the proposed model with signature sizes of $\{4, 8, 16, 32, 64, 128\}$. As seen in Table 1, we observe that our model is robust to varying sizes of the cardiovascular signatures, with a decrease in validation error that levels off after a size of 16.

Effect of training set size: Our model leverages a population to learn a single persons cardiovascular transfer function. To understand the effect of the population on the model, we vary the training set size as fractions of the total (1%, 5%, 10%, 50%, 100%) and observe how well our model performs. As seen in Table 1, we observe a steady decrease in validation error as the training data is increased, culminating in a 14% better performance with a full dataset as opposed to only 1% of it.

Signature consistency: To assess test-retest reliability, a measure of *internal validity*, we consider how well a cardiovascular signature can be used to predict a persons heart rate from their physical activity a year later. For each individual in the validation set, we learn a signature from their signal measurements during January 2017 and apply that signature to predict their heart rate during January 2018. As compared to using a different persons signature, a person’s own signature is significantly better at predicting their heart rate (Wilcoxon signed-rank test, $V = 2.6 \times 10^7$, $p < 10^{-16}$), with a median of 60% greater mean-square error when using another person’s, randomly selected.

Predicting health conditions using signatures: To assess the *external validity* of the signatures, we tested whether they are associated with factors affecting cardiovascular response, such as age and body mass index (BMI). We used an XGBoost model [6] trained on the learned size-32 validation signatures to predict if an individual is above/below median age of the cohort (31 years) with an AUC of 70.1% when trained on a random 70/30 split of the validation set. Predicting if a person is obese ($BMI \geq 30$) from solely their signature achieves an AUC of 69.7%. Predicting the same outcomes using only an individual’s resting heart rate results in significantly worse accuracy, with AUCs of 60.6% and 54.1%, respectively, demonstrating that signatures carry richer information about the relationship between physical activity and heart rate than the single RHR marker.

5 Discussion

It is informative to consider when a cardiovascular signature would *not* well predict a person’s heart rate. Assuming the measuring conditions of the wearable device stay the same, this may happen when a person’s cardiovascular response is hard to learn (e.g., short observation period, high missingness, or erratic behavior), when it changes (e.g., improvement/degradation of fitness), or when there are factors affecting HR that go beyond sleep and physical activity (e.g., stress endured during an interview, after taking medication, or having a meal). An example of where our model fails can be seen in the right-most column of Figure 1.

In future work we plan to explore the motifs surfaced by the attention component of the network, and study how they are related to health outcomes. From a methods perspective, future extensions will consider variational autoencoders to better condition the latent space of cardiovascular signatures as well as further hyper-parameter and architecture optimization.

References

- [1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.
- [2] Tim Althoff, Eric Horvitz, Ryan W White, and Jamie Zeitzer. Harnessing the web for population-scale physiological sensing: A case study of sleep and performance. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017.
- [3] Brandon Ballinger, Johnson Hsieh, Avesh Singh, Nimit Sohoni, Jack Wang, Geoffrey H. Tison, Gregory M. Marcus, Jose M. Sanchez, Carol Maguire, Jeffrey E. Olgin, and Mark J. Pletcher. Deepheart: Semi-supervised sequence learning for cardiovascular risk prediction. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018*, 2018.
- [4] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 2013.
- [5] G Brooks, T Fahey, and T White. Physiologic responses and long-term adaptations to exercise. *Exercise physiology: human bioenergetics and its applications*. 2nd ed. Mountain View (CA): Mayfield Publishing Co, pages 61–77, 1996.
- [6] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016.
- [7] François Chollet et al. Keras. <https://keras.io>, 2015.
- [8] Mandsager K, Harb S, Cremer P, Phelan D, Nissen SE, and Jaber W. Association of cardiorespiratory fitness with long-term mortality among adults undergoing exercise treadmill testing. *JAMA Network Open*, 1(6):e183605–, 2018.
- [9] DP Kingma and J Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
- [10] Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.
- [11] Jairo H Migueles, Cristina Cadenas-Sanchez, Ulf Ekelund, Christine Delisle Nyström, Jose Mora-Gonzalez, Marie Löf, Idoia Labayen, Jonatan R Ruiz, and Francisco B Ortega. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports medicine*, 47(9):1821–1845, 2017.
- [12] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010.
- [13] Mitesh S Patel, Luca Foschini, Gregory W Kurtzman, Jingsan Zhu, Wenli Wang, Charles AL Raeshide, and Susan M Zbikowski. Using wearable devices and smartphones to track physical activity: initial activation, sustained use, and step counts across sociodemographic characteristics in a national sample. *Annals of internal medicine*, 167(10):755–757, 2017.
- [14] Tom Quisel, Luca Foschini, Alessio Signorini, and David C Kale. Collecting and analyzing millions of mhealth data streams. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1971–1980. ACM, 2017.
- [15] Aäron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. In *SSW*, page 125, 2016.
- [16] A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, AN Gomez, Ł Kaiser, and I Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017.