Self-Supervised PPG Representation Learning Shows High Inter-Subject Variability

Ghorbani, Ramin; Reinders, Marcel J.T.; Tax, David M.J.

Citation (APA)
Ghorbani, R., Reinders, M. J. T., & Tax, D. M. J. (2023). Self-Supervised PPG Representation Learning Shows High Inter-Subject Variability. In *ICMLT 2023 - Proceedings of 2023 8th International Conference on Machine Learning Technologies* (pp. 127-132). (ACM International Conference Proceeding Series). Association for Computing Machinery (ACM). https://doi.org/10.1145/3589883.3589902

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Self-Supervised PPG Representation Learning Shows High Inter-Subject Variability

Ramin Ghorbani*
Pattern Recognition and
Bioinformatics group, Delft
University of Technology
Delft, Netherlands
r.ghorbani@tudelft.nl

Marcel J.T. Reinders
Pattern Recognition and
Bioinformatics group, Delft
University of Technology
Delft, Netherlands
m.j.t.reinders@tudelft.nl

David M.J. Tax
Pattern Recognition and
Bioinformatics group, Delft
University of Technology
Delft, Netherlands
D.M.J.Tax@tudelft.nl

ABSTRACT

With the progress of sensor technology in wearables, the collection and analysis of PPG signals are gaining more interest. Using Machine Learning, the cardiac rhythm corresponding to PPG signals can be used to predict different tasks such as activity recognition, sleep stage detection, or more general health status. However, supervised learning is often limited by the amount of available labeled data, which is typically expensive to obtain. To address this problem, we propose a Self-Supervised Learning (SSL) method with a pretext task of signal reconstruction to learn an informative generalized PPG representation. The performance of the proposed SSL framework is compared with two fully supervised baselines. The results show that in a very limited label data setting (10 samples per class or less), using SSL is beneficial, and a simple classifier trained on SSL-learned representations outperforms fully supervised deep neural networks. However, the results reveal that the SSL-learned representations are too focused on encoding the subjects. Unfortunately, there is high inter-subject variability in the SSL-learned representations, which makes working with this data more challenging when labeled data is scarce. The high inter-subject variability suggests that there is still room for improvements in learning representations. In general, the results suggest that SSL may pave the way for the broader use of machine learning models on PPG data in label-scarce regimes.

CCS CONCEPTS

• **Computing methodologies → Learning latent representations**.

KEYWORDS

Self-Supervised Learning, Representation Learning, Autoencoder, PPG Signal, Human Activity Recognition, Inter-Subject Variability

---

*Corresponding Author

1 INTRODUCTION

In recent years, wearables such as smartwatches and health trackers, equipped with a photoplethysmography (PPG) sensor, are becoming increasingly popular [7]. PPG is a non-invasive, low-cost optical measurement that can measure tissue blood flow over time following each pulse wave ejected from the heart. PPG works on the principle of pulse oximetry, wherein a sensor emits light to the skin and measures the intensity of light that is reflected or transmitted through the skin. Changes in arterial blood volume cause PPG signal variations [4, 14]. The cardiac rhythm corresponding to the PPG signal's periodicity can be used to obtain additional useful information from the users and predict various tasks. Some examples of research on PPG signals are related to Activity Recognition [6], Heart Rate Estimation [25], Blood Pressure Prediction [11], Biometric Identification [9], Sleep Staging Detection [23], and Atrial Fibrillation Detection [2].

In existing research, analyzing the PPG signals can be broadly categorized into signal processing and machine learning methods. The majority of machine learning solutions for PPG-based tasks utilize fully-supervised learning methods, which can be associated with several limitations. A fully-supervised learning setup usually requires considerable computational resources and time. Additionally, this setup requires large human-annotated datasets for high performance. Typically, obtaining labeled data is very costly and time-intensive, and the amount of labeled data is therefore insufficient in real-world applications, for instance, in the case of heart failure detection or fall detection. When automated detectors have to be trained on this type of problem, a good representation of the data with few numbers of informative features is essential [12]. Therefore, it is necessary to address the label-scarcity problem.

One approach to obtain a good informative representation is 'Self-Supervised Learning' (SSL). In SSL, two tasks are defined: a 'pretext' task and a 'downstream' task. The pretext task is the task of learning informative representations by itself. For instance, an auto-encoder tries to precisely reconstruct the input, squeezing the information through a bottleneck layer. It thereby learns a condensed, low-dimensional representation containing all necessary information to reconstruct the input exactly [22]. It is assumed that this learned

low-dimensional representation reduces the complexity of the data by reducing anomalies and noise and, at the same time, improves the ability to detect patterns in the data simpler and better. Hence, learned representations from the pretext task should be helpful for learning a second-stage classifier on the downstream task, which is the actual task of interest that we want to solve.

The latest research in the field of machine learning shows the potential of SSL for finding generalized and robust representations [5, 15, 21, 24]. The existing works in representation learning are generally concentrated on image-based applications where variations in the data could be visually observed. However, SSL is rarely applied to the field of time series data, especially biosignals. In recent years, some have applied SSL to time series data to show that this method can improve the representation, and they could confirm the potential of self-supervision in capturing important information even in the absence of labeled data. For instance, Lan et al. [17] introduced an Intra-inter Subject self-supervised Learning (ISL) model customized for ECG signals. Their model integrates medical knowledge into self-supervision to effectively learn from intra-inter subject differences. Their results over different evaluation scenarios showed that the learned representations are information-rich and more generalizable than other state-of-the-art methods for diagnosing cardiac arrhythmias in label-scarce regimes. As another example, Banville et al. [5] investigated SSL to learn representations from EEG signals. They explored two pretext tasks based on temporal context prediction and contrastive predictive coding on two clinically EEG-relevant downstream tasks. The results show that linear classifiers trained on SSL-learned representations consistently outperform purely supervised deep neural networks in label-scarce regimes while reaching competitive performance when all labels are available. These findings are, however, not yet shown on noisy PPG signals, so it still remains to be shown whether self-supervision can bring improvements over standard supervised approaches on PPG signals.

In this paper, we focus on Human Activity Recognition (HAR) from PPG data. This is gaining interest since PPG data can be easily acquired from any of the widely available wearable devices [3]. Researchers have been exploring how SSL techniques can be either extended or explicitly designed for HAR tasks on accelerometer and gyroscope data. However, they have not yet looked into the PPG data specifically. In one of the early pioneering works, Saeed et al. [19] used the task of identifying which signal transformation has been applied to a particular data sample as a pretext task using accelerometer and gyroscope data. The results show that SSL drastically reduces the requirement of labeled activity data, narrowing the gap between supervised and unsupervised techniques for learning meaningful representations.

Concluding, to the best of our knowledge, there are currently no studies using SSL on PPG data in label-scarce regimes. Therefore, we present the first detailed analysis of SSL tasks on PPG signals with attention to Activity Recognition as a downstream task. Our main contributions are: 1) Proposing a SSL framework for PPG data in label-scarce regimes, 2) Evaluating whether human activity recognition task can be done better when using SSL representations, and 3) Investigating the Inter-subject variability in PPG data and exploring how this is captured by the SSL representation.

2 PROPOSED FRAMEWORK

An overview of the proposed SSL framework is shown in Figure 1. We use an Autoencoder (AE) to learn a representation of the (unlabeled) data (unsupervised learning). Given an unlabeled dataset $D_U = \{\mathbf{x}_i\}_{i=1}^{N_u}$ where $\mathbf{x}_i \in \mathbb{R}^{1 \times T}$ is a vector of length T and N_u is the number of vectors (samples). The encoder maps each input vector into a latent space representation $\mathbf{h}_i = E_\phi(\mathbf{x}_i)$ where $\mathbf{h}_i \in \mathbb{R}^{1 \times d}$ where $d < T$. After that, $\mathbf{h}_i$ is fed into the decoder component of the model, which follows the same approach to map $\mathbf{h}_i$ to the output values $\hat{\mathbf{x}}_i = D_\theta(\mathbf{h}_i)$ where $\hat{\mathbf{x}}_i \in \mathbb{R}^{1 \times T}$. The encoder and decoder are parametrized by ϕ and θ, respectively. The AE is trained to minimize the mean squared error between $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ [16]:

$$\mathbf{L}_{Total} = 1/N_u \sum_{i=1}^{N_u} (1/T \parallel \mathbf{x}_i - \hat{\mathbf{x}}_i \parallel^2) \qquad (1)$$

We have used a combination of Convolutional Neural Network layers (CNN) with the AE architecture, a Convolutional Neural Network AutoEncoder (CNN-AE). The main advantage of using a CNN-AE is a better reconstruction for the PPG signal as it exploits the correlations between time measurements in the PPG signal and thus captures the time-dependent information better.

For the downstream task, we use the original preprocessed PPG as the input to E_ϕ with frozen trained weights from the reconstruction task to get the related latent representation $\mathbf{h}_i$. This $\mathbf{h}_i$ is used as input to train a simple classifier such as Logistic Regression (LR) or k-Nearest Neighbors (kNN). We do this because when the number of training samples is small, these simple classifiers often outperform more flexible and complex models. The classifier is trained on $\mathbf{h}_i$'s from several subjects and tested on a completely new subject.

3 EXPERIMENTAL SETUP

3.1 Dataset

We use the PPG-Dalia public dataset, which was collected by Reiss et al. [18] for the PPG-based heart rate estimation task. This dataset contains recordings of 15 subjects performing daily activities such as *sitting, ascending/descending stairs, playing table soccer, cycling, driving a car, having lunch, walking,* and *working*. The measurements are obtained from wrist and chest-worn devices. Besides the activities, the transient periods between the activities are also recorded. We removed data from subject number 6 due to hardware issues during data recording. Note that *having lunch, driving a car*, and *working* activities are categorized as concurrent or inter-leaved human activities, where actions of multiple activities are carried out simultaneously or where activities contain various activities while their actions can be interleaved in a shuffled manner [1, 20]. Therefore, we only considered the remaining five human activities for further study.

3.2 Data Preprocessing

For the pretext task, a band-pass 2th order Butterworth filter with low and high frequencies of 0.1 - 6 Hz is applied to the whole PPG signal of each subject individually. The filtered signal is normalized to zero mean and unit variance per subject. The final normalized filtered signals are segmented into a fixed window size of 8 seconds
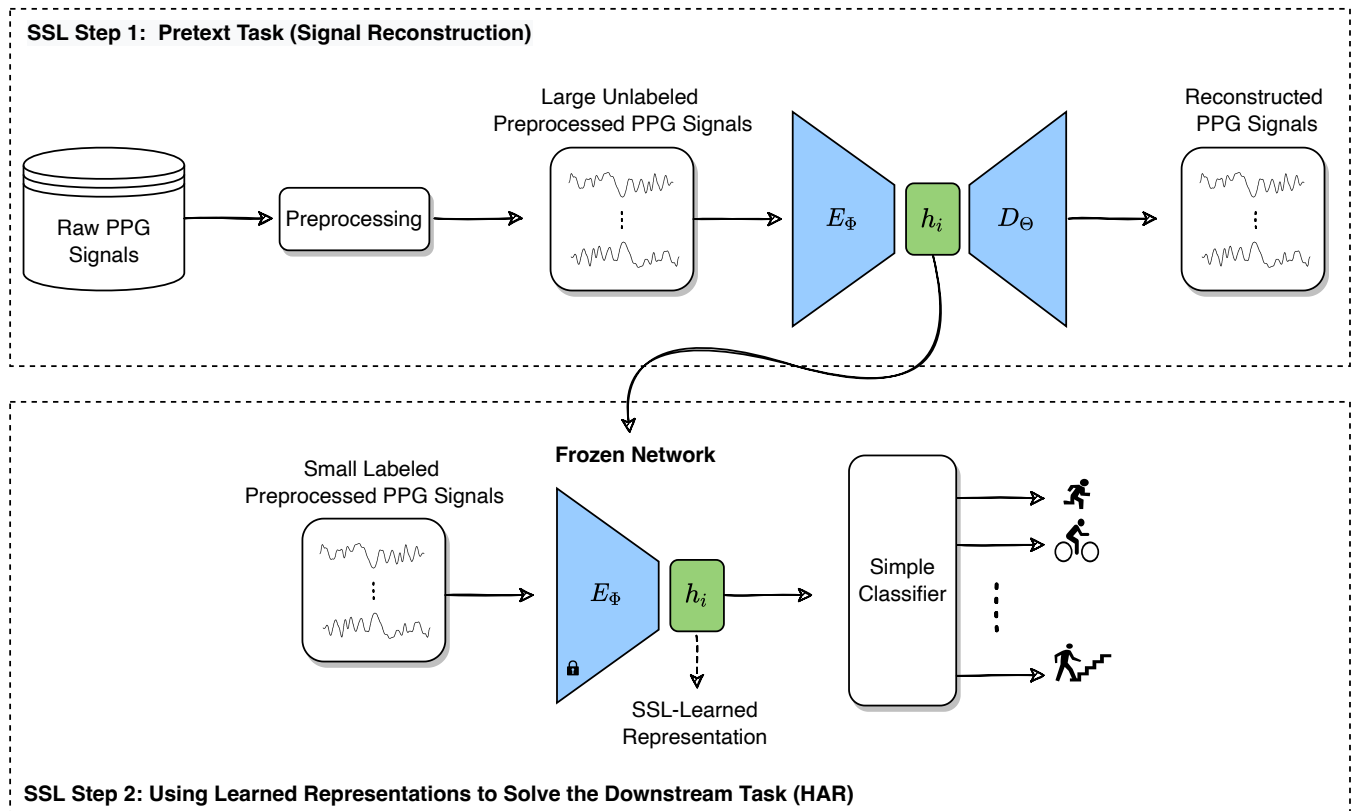
Figure 1: The overall proposed Self-Supervised framework

while two successive windows overlap by 6 seconds in the test dataset (this setting is common for PPG data). To increase the training set size, the two successive windows overlap by 7 seconds in the training dataset. For the downstream task, the PPG signals are split into 8 seconds windows while two successive windows overlap by 6 seconds in both training and test datasets. Activity labels are assigned to the corresponding PPG windows based on the available annotations.

3.3 Implementation

3.3.1 Proposed SSL. The hyperparameters and the architecture of the proposed CNN-AE are systematically determined by searching through all possible combinations to obtain the best performance. Eventually, we used a CNN-AE architecture deep learning model consisting of three convolution layers, followed by the Exponential Linear Unit (ELU) activation function, Batch Normalization, and MaxPooling layers. The decoder consists of the hidden layers in the reverse order of the encoder section. The Adam optimizer with a learning rate of 0.01, a decay rate of 0.001, and a clip-norm value of 0.9 are used. The batch size is 128, and training runs for 200 epochs. Finally, the parameters of all layers are randomly initialized. To assess the randomness of the deep learning framework, each training process for each test subject is repeated five times. Leave-One-Subject-Out cross-validation (LOSO) is used to evaluate the reconstruction performance.

For the downstream task, two simple classifiers are trained on the SSL-learned representations separately: a Logistic Regression and a kNN classifier (SSL-LR and SSL-kNN, respectively). The SSL-LR is regularized with the L2 penalty term and is solved using LIBLIN-EAR [10]. The SSL-kNN is trained with reweighted neighbors [8], where points are weighted by the inverse of their distance. Therefore, closer neighbors of a query point will have a larger influence than far away neighbors. Due to the different number of training samples which are 2, 5, 10, 50, and 1000 per class, the number of neighbors is selected as 8, 19, 39, 115, and 350, respectively. The LOSO is used to evaluate the AUC performance.

3.3.2 Comparative Baselines. The performance of the SSL method is compared with two other baseline models: a simple and a more complex one. The simple baseline model (a typical baseline in SSL research) is trained directly on the original preprocessed PPG representations and consists of the encoder part of the CNN-AE from the pretext task, extended with one classification layer at the end. The encoder part is thus trained on the classification task immediately and not in a self-supervised setting. The Adam optimizer with a learning rate of 0.001 and a clip-norm value of 0.6 are used. The batch size is 128, and training runs for 200 epochs. The more complex baseline is a CNN-LSTM model also trained directly on the original preprocessed PPG representations. The architecture of the complex baseline consists of a convolution layer followed by a hyperbolic tangent function, Batch Normalization, MaxPooling

Table 1: Mean Relative MSE results of test subjects (LOSO) for signal reconstruction task by varying d using the CNN-AE

Dimension of the h_i	Relative MSE Results
$d = 2$	0.83 ± 0.02
$d = 8$	0.59 ± 0.03
$d = 32$	0.14 ± 0.03
$\boldsymbol{d = 64}$	$\boldsymbol{0.02 \pm 0.00}$
$d = 128$	0.00 ± 0.00

layers, and then a LSTM layer with a hyperbolic tangent activation function, followed by a classification layer at the end. The Adam optimizer with a learning rate of 0.001 and a clip-norm value of 0.6 are used. The batch size is 128, and training runs for 200 epochs. To assess the randomness of these Deep Learning frameworks, each training process for each test subject is repeated five times. Both baseline models use the LOSO to evaluate the AUC performance.

3.3.3 Biometric Identification (BI) for Exploring Inter-Subject Variability. If there is a large inter-subject variability, the subjects should be easily discriminated in the representation. To check if a subject can indeed be easily discriminated, we train and evaluate a kNN classifier with reweighted neighbors ($k = 20$) per activity. Note that this classifier is not optimized at all on the SSL-learned representations; the kNN fully relies on the metric that is induced by the CNN-AE latent representation h_i. A good performance of the kNN for the BI task suggests that the learned representation from CNN-AE is heavily biased towards encoding different subjects and not so many other tasks like activities. In this experiment, PPG data is preprocessed with the same steps as the downstream task preprocessing. Afterward, the PPG windows of each activity are selected to sample a separate balanced training set over the subjects. The 4-fold cross-validation (75% for the training and 25% for the test set) is used to evaluate the AUC performance.

4 RESULTS

4.1 Pretext Task

To determine a suitable dimensionality d of the learned representation h_i, we compute the relative MSE (i.e., MSE in Eq. (1) normalized by the total variance across test subjects' data) by varying d using the CNN-AE. The results are shown in Table 1. It can be seen that the reconstruction error decreases with increasing dimensionality d. As the representation with a lower dimension is more suited for learning with limited labels, we chose $d = 64$ when proceeding with the downstream task. Also, later experiments show that $d = 64$ leads to better performance on the downstream task compared to other dimensionalities.

4.2 Downstream Task

In Figure 2a, we show the AUC performances on the downstream task of predicting activity type for a varying number of training samples per class. The performance of the proposed SSL method is compared with the simple and complex baseline methods. As the number of training samples per class decreases, the performances of all methods drop, confirming the negative influence of when

less and less samples with labels are available. The linear SSL-LR model fails to improve the performance compared to the baseline models when a few (< 10) training samples per class are available. However, SSL-kNN, as a non-linear solution, outperforms the baselines and the SSL-LR in the label-scarce regimes. This suggests that the SSL-learned representation is still too complex for a simple linear solution like LR. One reason for such a behavior could be the high Inter-Subject variability in the SSL-learned representations. Figure 2b shows the SSL-kNN performance of each of the individual test subjects, for a varying number of training samples per class. It can be observed that the AUC performance can vary between 0.5 and 0.7 for a small training size of $N = 2$, and even for very large training sizes of $N = 1000$, the AUC still varies between 0.55 and 0.8. This indicates that the data distributions of different subjects vary significantly, thus indicating large inter-subject variability.

To explore the inter-subject variability more deeply, we also investigated the possibility of classifying subjects during each activity using a kNN model. Note that this is now different than our initial domain task; here, we are interested in whether subjects are still separable in the latent representations h_i (which, in principle, is undesired when generalizing over subjects). The results of the BI task in Table 2 show that subjects can be discriminated perfectly for some activities such as *sitting* or *walking*. It can be seen that there are enough differences among subjects in the original as well as the SSL-learned representations. Moreover, the SSL-learned representation seems to highlight the inter-subject variation as its mean performance is consistently higher than on the original data. This suggests that the AE learned representation is more focused on encoding the different subjects and not so much on the domain tasks of interest, that is, predicting the different activities.

5 DISCUSSION AND CONCLUSION

We have evaluated the usefulness of self-supervised representation for the activity recognition task when suffering from a label-scarcity in PPG data. The representation is not optimized on the downstream classification task (for which just a few labeled training samples may be available), but it is first optimized to perform a data reconstruction pretext task (for which no supervised information is needed). The results reveal that the SSL method can compete and outperform fully supervised baselines when a kNN model is trained on the SSL-learned representations in label-scarce regimes (with less than 50 samples per class). However, training a simple linear classifier like LR (instead of kNN) is not helpful since the inter-subject variability introduces too much non-linearities in the decision boundaries.

One should note that in the current study setup, fixed hyperparameters are used across the data regimes for all baseline models. When copious amounts of (unlabeled) data from all subjects would be available, all hyperparameters could be optimized for every different task.

The poor performance of the LR classifier on the SSL-learned representations shows that there is high inter-subject variability. High inter-subject variability makes the generalization more challenging. In this case, a subject-specific model could be a solution for

Table 2: Mean AUC Performance of test sets (4-fold cross-validation) for Biometric Identification task

Activities	Input	
	Original Representation	SSL-Learned Representation
Sitting	0.84 ± 0.12	0.84 ± 0.01
Ascending/Descending Stairs	0.63 ± 0.02	0.64 ± 0.03
Playing Table Soccer	0.53 ± 0.01	0.61 ± 0.02
Cycling	0.59 ± 0.05	0.61 ± 0.06
Walking	0.72 ± 0.02	0.75 ± 0.02
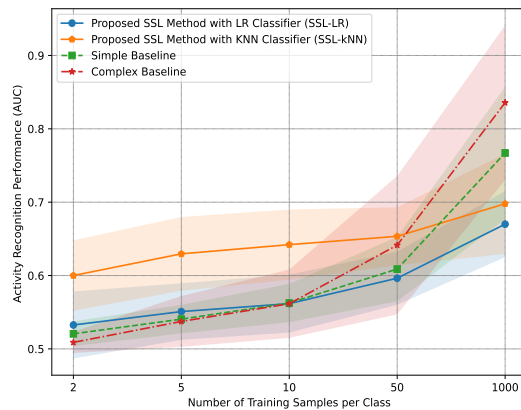
(a)
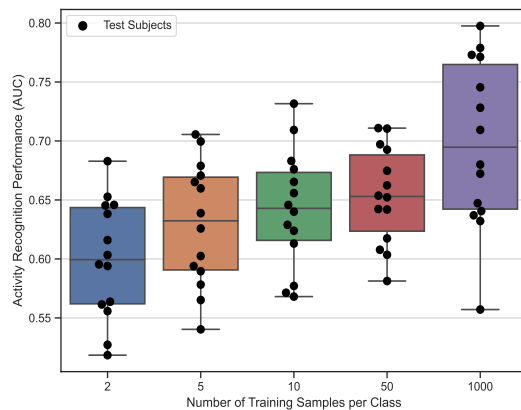
(b)

Figure 2: a) Mean AUC performance of test subjects (using LOSO) for Activity Recognition downstream task over the different number of training samples per class. b) The variability among different test subjects in AUC performance for Activity Recognition downstream task using SSL-kNN

improving the performance over the learned representations. However, training a subject-specific model can be expensive because a large amount of (labeled) data that has to be obtained from each subject. Here, the SSL representation can come to the rescue, as we

have shown that this representation can improve performance with respect to the original representation. However, there should be more focus on disentangling the inter- and intra-subject variability.

This matter opens the door for future research to learn more generalized informative PPG representations while addressing the inter-subject variability problem. For instance, removing the subject-specific factors in order to disentangle the inter-subject variations using a factor disentangling sequential autoencoder [13], or performing contrastive learning among subjects to learn distinctive representations [17] can be promising directions in learning informative PPG representations.

REFERENCES

[1] Mahsa Sadat Afzali Arani, Diego Elias Costa, and Emad Shihab. 2021. Human Activity Recognition: A Comparative Study to Assess the Contribution Level of Accelerometer, ECG, and PPG Signals. *Sensors* 21, 21 (2021), 6997.

[2] Alireza Aliamiri and Yichen Shen. 2018. Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, IEEE, Las Vegas, NV, USA, 442–445.

[3] Omair Rashed Abdulwareth Almanifi, Ismail Mohd Khairuddin, Mohd Azraai Mohd Razman, Rabiu Muazu Musa, and Anwar PP Abdul Majeed. 2022. Human activity recognition based on wrist PPG via the ensemble method. *ICT Express* (2022).

[4] Takuo Aoyagi and Katsuyuki Miyasaka. 2002. Pulse oximetry: its invention, contribution to medicine, and future tasks. *Anesthesia and analgesia* 94, 1 (2002), S1–S3.

[5] Hubert Banville, Omar Chehab, Aapo Hyvärinen, Denis-Alexander Engemann, and Alexandre Gramfort. 2021. Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering* 18, 4 (2021), 046020.

[6] Mehdi Boukhechba, Lihua Cai, Congyu Wu, and Laura E Barnes. 2019. ActiPPG: using deep neural networks for activity recognition from wrist-worn photoplethysmography (PPG) sensors. *Smart Health* 14 (2019), 100082.

[7] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, Cinna Soltanpur, and Homer Nazeran. 2018. A review on wearable photoplethysmography sensors and their potential future applications in health care. *International journal of biosensors & bioelectronics* 4, 4 (2018), 195.

[8] Sahibsingh A Dudani. 1976. The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics* 4 (1976), 325–327.

[9] Luke Everson, Dwaipayan Biswas, Madhuri Panwar, Dimitrios Rodopoulos, Amit Acharyya, Chris H Kim, Chris Van Hoof, Mario Konijnenburg, and Nick Van Helleputte. 2018. BiometricNet: Deep learning based biometric identification using wrist-worn PPG. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, IEEE, 1–5.

[10] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *the Journal of machine Learning research* 9 (2008), 1871–1874.

[11] Shrimanti Ghosh, Ankur Banerjee, Nilanjan Ray, Peter W Wood, Pierre Boulanger, and Raj Padwal. 2016. Continuous blood pressure prediction from pulse transit time using ECG and PPG signals. In *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*. IEEE, IEEE, 188–191.

[12] Prashnna Kumar Gyawali. 2021. *Learning with Limited Labeled Data in Biomedical Domain by Disentanglement and Semi-Supervised Learning*. Rochester Institute of Technology.

[13] Prashnna Kumar Gyawali, B Milan Horacek, John L Sapp, and Linwei Wang. 2019. Sequential factorized autoencoder for localizing the origin of ventricular activation from 12-lead electrocardiograms. *IEEE Transactions on Biomedical Engineering* 67, 5 (2019), 1505–1516.

[14] AAR Kamal, JB Harness, G Irving, and AJ Mearns. 1989. Skin photoplethysmography—a review. *Computer methods and programs in biomedicine* 28, 4 (1989), 257–269.

[15] Muhammed Kocabas, Salih Karagoz, and Emre Akbas. 2019. Self-supervised learning of 3d human pose using multi-view geometry. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1077–1086.

[16] Mark A Kramer. 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal* 37, 2 (1991), 233–243.

[17] Xiang Lan, Dianwen Ng, Shenda Hong, and Mengling Feng. 2022. Intra-inter subject self-supervised learning for multivariate cardiac signals. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 4532–4540.

[18] Attila Reiss, Ina Indlekofer, Philip Schmidt, and Kristof Van Laerhoven. 2019. Deep PPG: Large-scale heart rate estimation with convolutional neural networks. *Sensors* 19, 14 (2019), 3079.

[19] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. 2019. Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–30.

[20] Nehal A Sakr, Mervat Abu-Elkheir, Ahmed Atwan, and HH Soliman. 2019. Data driven recognition of interleaved and concurrent human activities with nonlinear characteristics. *Journal of Intelligent & Fuzzy Systems* 37, 4 (2019), 5573–5588.

[21] Marco Tagliasacchi, Beat Gfeller, Félix de Chaumont Quitry, and Dominik Roblek. 2019. Self-supervised audio representation learning for mobile devices. *arXiv preprint arXiv:1905.11796* (2019).

[22] Naftali Tishby, Fernando C Pereira, and William Bialek. 2000. The information bottleneck method. *arXiv preprint physics/0004057* (2000).

[23] Muhammed Kürşad Uçar, Mehmet Recep Bozkurt, Cahit Bilgin, and Kemal Polat. 2018. Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques. *Neural Computing and Applications* 29, 8 (2018), 1–16.

[24] Jiangliu Wang, Jianbo Jiao, Linchao Bao, Shengfeng He, Yunhui Liu, and Wei Liu. 2019. Self-supervised spatio-temporal representation learning for videos by predicting motion and appearance statistics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4006–4015.

[25] Zhilin Zhang, Zhouyue Pi, and Benyuan Liu. 2014. TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on biomedical engineering* 62, 2 (2014), 522–531.