Personalized anomaly detection in PPG data using representation learning and biometric identification

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

DOI
10.1016/j.bspc.2024.106216

Publication date
2024

Document Version
Final published version

Published in
Biomedical Signal Processing and Control

Biomedical Signal Processing and Control 94 (2024) 106216

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Personalized anomaly detection in PPG data using representation learning and biometric identification

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

Pattern Recognition and Bioinformatics group, Delft University of Technology, Delft, Netherlands

ARTICLE INFO

Keywords:
Representation learning
PPG
Anomaly detection
Multivariate normal distribution
Isolation forest
PCA reconstruction

ABSTRACT

Photoplethysmography (PPG) signals, typically acquired from wearable devices, hold significant potential for continuous fitness-health monitoring. In particular, heart conditions that manifest in rare and subtle deviating heart patterns may be interesting. However, robust and reliable anomaly detection within these data remains a challenge due to the scarcity of labeled data and high inter-subject variability. This paper introduces a two-stage framework leveraging representation learning and personalization to improve anomaly detection performance in PPG data. The proposed framework first employs representation learning to transform the original PPG signals into a more discriminative and compact representation. We then apply three different unsupervised anomaly detection methods for movement detection and biometric identification. We validate our approach using two different datasets in both generalized and personalized scenarios. Our results demonstrate significant improvements: for movement detection, in the generalized scenario, AUCs improved from barely 0.5 to above 0.9 with representation learning. Importantly, inter-subject variability was substantially reduced, from around 0.4 to below 0.1. In the personalized scenario, AUCs became close to 1.0, with variability further reduced to below 0.05, indicating the effectiveness of both representation learning and personalization for anomaly detection in PPG data. Similar enhancements were observed in biometric identification, emphasizing how our approach can minimize inter-subject variability and enhance PPG-based health monitoring systems.

1. Introduction

Photoplethysmography (PPG) data is a non-invasive, low-cost, optical physiological signal that measures the volume of blood flowing through the blood vessels and can be measured by a variety of wearable devices and smartwatches [1]. PPG data enables remote health monitoring and fitness tracking, which presents opportunities for identifying unusual patterns in the user data that may indicate potential health issues, like abnormal heart rate or irregular movement patterns [2].

The effectiveness of detecting anomalies largely depends on the availability of enough labeled data. Supervised machine learning methods, such as k-Nearest Neighbors (kNN), Random Forest, and Artificial Neural Networks (ANN), have been widely used in previous research to interpret PPG signals [3–7]. However, the process of data labeling is tedious, time-consuming, and costly, especially for anomaly detection problems, since anomalies seldomly occur in real-world applications. Furthermore, these supervised learning methods may be prone to bias and overfitting if the labeled dataset does not adequately represent the full range of normal and anomalous PPG signals. Moreover, these methods may not be adaptable to unknown or unexpected anomalies, as they learn to recognize patterns based on the examples provided in the training dataset, and may fail to effectively detect anomalies not represented in the dataset. To address these limitations, unsupervised anomaly detection methods can offer advantages over supervised approaches, as they do not rely on explicitly labeled examples of anomalous behavior and can be more adaptable to unknown or unexpected anomalies [8,9].

Anomaly detection in PPG data can also be challenging due to other various factors that contribute to noise and inter-subject variability. These are factors like physical activity, stress, illness, measurement noise, age, gender, body composition, and genetic differences, as well as external factors such as sensor placement, sensor quality, and environmental conditions. This makes it difficult to develop generalized models that perform consistently across different individuals since each person's PPG signal may exhibit unique characteristics [10]. These complexities necessitate strategies to account for individual-specific characteristics.

Personalization can be a potential solution to help overcome the limitations of generalization by tailoring models to individual users [11]. However, the effectiveness of personalization hinges on accurate biometric identification. Inaccurate identification of individuals can

* Corresponding author.
E-mail address: r.ghorbani@tudelft.nl (R. Ghorbani).

https://doi.org/10.1016/j.bspc.2024.106216
Received 14 July 2023; Received in revised form 2 February 2024; Accepted 9 March 2024
Available online 20 March 2024

lead to personalized models being trained on or applied to the wrong user's data, resulting in poor performance and potentially harmful outcomes. Hence, accurate biometric identification can enhance the reliability of personalized models, as it ensures that the models are based on the specific characteristics of each user.

In addition to unsupervised anomaly detection methods and personalization, representation learning can be particularly useful in enhancing performance [12]. Representation learning models are typically trained to learn from large amounts of unlabeled data, enabling them to extract a more compact, informative, and expressive representation of the data without the need for expensive and time-consuming manual labeling. By learning a lower-dimensional representation of the PPG data that captures its inherent structure and discriminative features, representation learning can help overcome challenges posed by inter-subject variability, noise, and other factors affecting PPG signals. AutoEncoders, for example, are a type of self-supervised representation learning model that learns representations by encoding inputs into lower dimensions and then decoding them back to their original form, focusing on reconstructing the input [13]. Other representation learning models have been proposed with different tasks, such as contrastive learning or classification of augmented transformations of the original data [14,15]. Representation learning has been increasingly used for anomaly detection in various domains, including image analysis [16–20], and time series data, such as bio-signals sensor data like EEG or ECG [21–24]. These studies have shown how representation learning can successfully extract meaningful features from complex bio-signals sensor data, leading to improved performance in classification tasks such as emotion detection or sleep stage classification. However, its application to PPG data for unsupervised anomaly detection and biometric identification remains underexplored, despite PPG being a commonly used bio-signal in health-monitoring applications.

In this paper, we present a two-stage framework for unsupervised movement detection and biometric identification in PPG data using representation learning. In the first stage, we train a deep neural network to obtain a lower-dimensional and informative data representation. In the second stage, we construct separate unsupervised anomaly detectors for both tasks using the learned representations from the first stage. Our approach not only investigates the effectiveness of representation learning in this context, but also explores the potential of personalization in enhancing anomaly detection performance. Additionally, we delve into biometric identification, aiming to improve the reliability of personalized anomaly detectors. To the best of our knowledge, this is the first study to jointly address these aspects for anomaly detection in PPG data. Summarizing, our contributions are:

1. We propose a two-stage framework for unsupervised anomaly detection and biometric identification in PPG data using representation learning.
2. We demonstrate the effectiveness of using the learned representations compared to the original representations in detecting difficult real-world anomalies and mitigating the subject variability.
3. We compare the effectiveness of generalization and personalization in anomaly detection, discussing the impact of tailoring models to individual users for enhancing the detection performance.
4. We investigate the unsupervised biometric identification task in PPG data to increase the reliability of personalized models.
5. We explore the impact of the dimensionality of the learned representation on the performance of our anomaly detection framework, demonstrating the robustness of representation learning across a wide range of dimensionalities.

2. Proposed framework

An overview of the proposed anomaly detection framework is shown in Fig. 1. In the first step, we focus on obtaining a representation of the PPG data that captures the underlying structure of the data. Recent research shows that the task of classifying the original data and augmented transformed versions of the same data can outperform AutoEncoders and contrastive learning methods in learning better representation for the downstream task of interest [15]. Accordingly, we learn the representation by distinguishing original data from augmented transformed versions of the same data. This task is what we refer to as "Signal Transformation Classification".

Given the original signal $S(l)$, where $l = (1, 2, \dots, L)$ and L is the length of the time series, the augmented transformations of the data are described as:

- *Time reversal*: A time inverted version of the signal: $S'(l)$, where $l = (L, L-1, \dots, 1)$.
- *Amplitude reversal*: A amplitude inverted version of the signal: as $S'(l) = -S(l)$, where $l = (1, 2, \dots, L)$.
- *Both Time and Amplitude reversal*: We first perform the time reversal as described and then perform the amplitude reversal to obtain a time and amplitude inverted version of the signal: $S'(l) = -S(l)$, where $l = (L, L-1, \dots, 1)$.

To train the representations, we use a CNN model to classify PPG segments into four categories: Time reversal, Amplitude reversal, Both time and amplitude reversal, and the original signal. Given an unlabeled PPG 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). $y_i \in \{1, 2, 3, 4\}$ is the class label for the ith vector, where $y_i = 1, 2, 3$ represents the augmented data obtained by reversing the original PPG signal and $y_i = 4$ represents the original PPG signal. The CNN model consists of an encoder component that maps each input vector into a latent space representation $\mathbf{h}_i = E_\phi(\mathbf{x}_i)$ where $\mathbf{h}_i \in \mathbb{R}^{1\times d}$ and $d < T$. After that, $\mathbf{h}_i$ is fed into the classifier component of the model to predict the class label $\hat{y}_i = C_\theta(\mathbf{h}_i)$. The model is trained to minimize the cross-entropy loss between the predicted class label $\hat{y}_i$ and the true class label y_i. The final learned representation is obtained by taking the latent space representation $\mathbf{h}_i$ outputted by the encoder component. This learned representation is then used in the second stage of our proposed framework for anomaly detection.

In the second step of our proposed framework, we use the learned representation $\mathbf{h}_i$ to detect anomalies. Specifically, we use three different methods to detect whether an input signal is an anomaly: Multi-Variate Normal distribution (MVN) [25], Isolation Forest (IF) [26], and PCA-Reconstruction [25]. For the MVN, the mean and covariance matrix are estimated on normal training samples. Given a test sample $\mathbf{h}_{test}$, we can then calculate the probability density function (PDF) of the test sample using the fitted Gaussian distribution as:

$$p(\mathbf{h}_{test}) = \frac{1}{(2\boldsymbol{\pi})^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{h}_{test} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{h}_{test} - \boldsymbol{\mu})\right) \quad (1)$$

where d is the dimension of the learned representation. It is expected that anomalous test samples have a lower probability compared to normal samples. Therefore, these points can be detected if the probability is below a set threshold.

For the Isolation Forest (IF) method, we first train an ensemble of decision trees on normal training samples. Given a test sample $\mathbf{h}_{test}$, the IF algorithm isolates the test sample from the others by recursively splitting the data with randomly selected features and split values. The number of splits, or the path length, required to isolate a sample is an indication of its anomaly score. Anomalous samples are expected to have shorter path lengths compared to normal samples. The anomaly score of a test sample $\mathbf{h}_{test}$ using the IF algorithm is calculated as:

$$s(\mathbf{h}_{test}) = 2^{-\frac{E[L(\mathbf{h}_{test})]}{c(N)}} \quad (2)$$

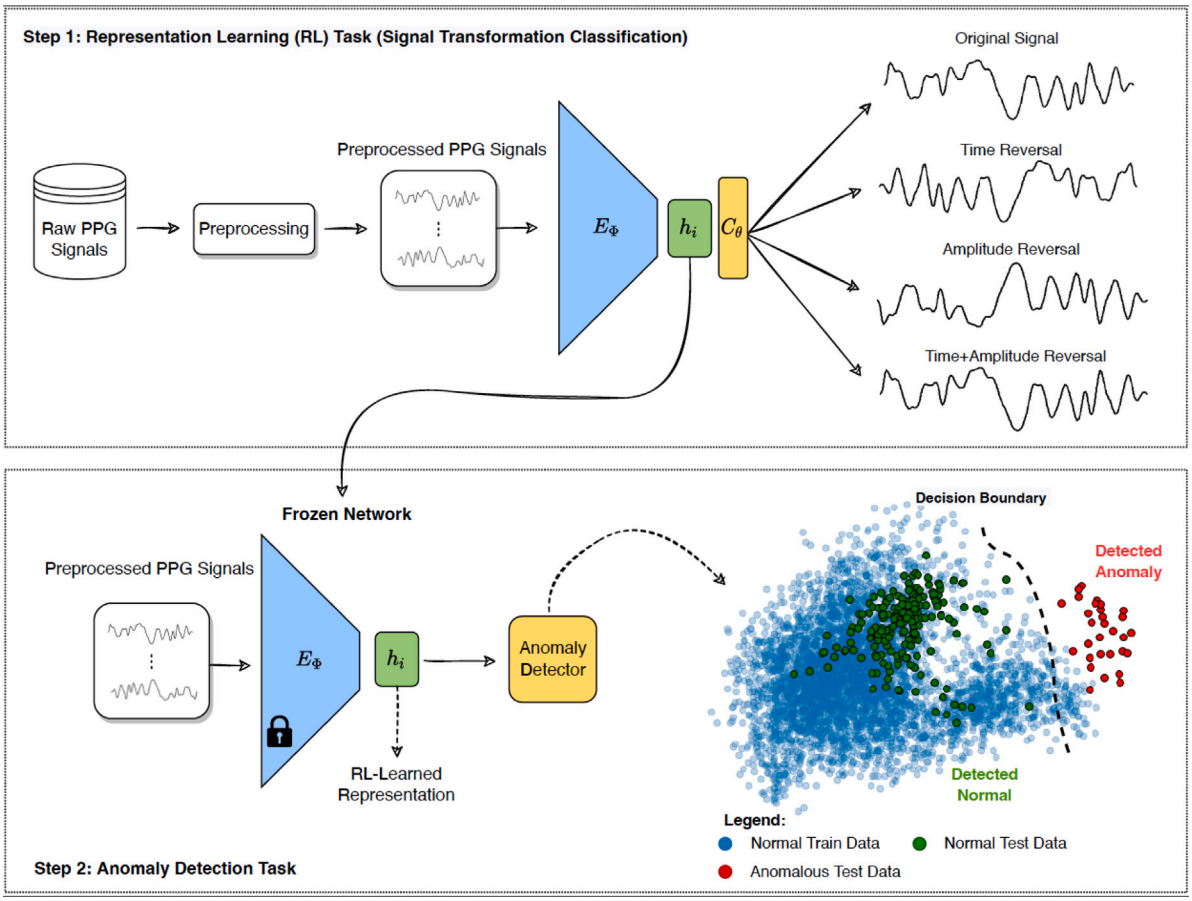

Fig. 1. *Proposed framework for anomaly detection using Representation Learning (RL)*. The framework consists of two steps: (1) A representation learning phase, where the model (consisting of an encoder and classifier component) is trained to discriminate between augmented transformations and the original data. The weights of the encoder are then frozen for the next step. (2) A anomaly detection phase, where the frozen encoder is used to extract features from the input data, which are then fed into an anomaly detector. The scatter plot illustrates an example of the distribution of points in a 2-dimensional feature space. The anomaly detector separates the normal and anomalous samples with the decision boundary (threshold) based on their anomaly scores.

where $E[L(\mathbf{h}_{test})]$ is the average path length of the test sample over all trees in the ensemble, $c(N)$ is the average path length of an unsuccessful search in a Binary Search Tree with N external nodes, and N is the number of samples in the training data. The anomaly score $s(\mathbf{h}_{test})$ ranges from 0 to 1, with higher scores indicating a higher likelihood of being anomalous. Anomalous test samples can be detected if the anomaly score is above a set threshold.

The PCA-Reconstruction method is a technique for detecting anomalies in high-dimensional data by reconstructing the original data from its principal components and evaluating the reconstruction error. Given a test sample $\mathbf{h}_{test}$, the reconstruction error can be calculated as the squared distance between the original sample and its reconstructed version ($\mathbf{h}_{recon}$) after mapping to a reduced PCA space. This is achieved by projecting the test sample $\mathbf{h}_{test}$ onto the orthogonal basis vectors represented by the matrix describing the PCA mapping, W, and then transforming it back to the original space. The reconstruction error can then be expressed as:

$$e(\mathbf{h}_{test}) = \|\mathbf{h}_{test} - (WW^T)\mathbf{h}_{recon}\|^2 \tag{3}$$

Note that anomalous test samples can be detected if the reconstruction error is above a set threshold.

2.1. Definition of anomalies

We define anomalies in the context of two specific tasks: activity movement detection and biometric identification.

2.1.1. Activity movement detection

In this particular setting, we train an anomaly detector on the recorded data during a specific activity (considered as the "normal" activity) and evaluate it on the data, which includes another activity (considered as an "anomalous" activity) in addition to the "normal" activity. We assume that the anomalous movement activity shows a different pattern than the normal activity and should be distinguishable from the "normal" movement activity. Accurately detecting movement can have significant practical implications in various applications, such as fitness health tracking, where identifying irregular patterns or deviations from expected behavior is crucial. By focusing on such a complex and practical problem, we can demonstrate the effectiveness and robustness of our proposed approach in handling real-world challenges associated with PPG data, including inter-subject variability, noise, and other factors affecting signal quality.

2.1.2. Biometric identification

In the context of biometric identification, we aim to identify an individual (user) as an anomaly when compared to a given group of people or another individual as the "intended" user(s). We train the anomaly detector on the recorded data from the intended group or individual during a specific activity and evaluate the anomaly detector when presenting new data, which includes another individual (considered as an "anomaly") during the same activity as the data from the intended user(s). Identifying such anomalies can be crucial in personalized health monitoring systems, where it is important to distinguish between users for accurate and safe health monitoring and assessments.

3. Experimental setup

3.1. Datasets

We use two datasets in our experiments. The first one is the Pulse Transit Time PPG (PTT-PPG) public dataset [27], a high-resolution and time-synchronized dataset annotated with activity labels. It contains waveform records from multi-wavelength sensors measuring PPGs, attachment pressures, and temperatures. The recordings are from 22 healthy subjects (M = 22) performing different physical activities in random order. We selected *Sitting* and *Walking* activities for this study. We use the green wavelength recorded PPG from the proximal phalanx (base segment) of the left index finger palmar side (Frequency of 500 Hz).

The second dataset is the PPG-Dalia public dataset collected by Reiss et al. [28] to perform PPG-based heart rate estimation. It has recordings of 15 subjects (M = 15) performing different daily activities. We have selected *Sitting* and *Walking* activities for this study. We removed data from subject number 6 due to incomplete data recording. The signals are recorded with a frequency of 64 Hz.

3.2. Data preprocessing

A band-pass 2nd order Butterworth filter is applied to the whole PPG signal of each subject individually for both datasets, but with different frequency ranges of 0.35–20 Hz for the PTT-PPG dataset and 0.1–10 Hz for PPG-Dalia. To create different categories of signals, we used Time reversal, Amplitude reversal, and both Time and Amplitude reversal augmentations. All of the signals are then normalized to zero mean and unit variance across the whole signal per subject. The final normalized filtered signals are segmented into windows with a length of 8 s, while two successive windows overlap by 7.5 s (this setting is common for PPG data [28–30]). Since the PTT-PPG dataset frequency is 500 Hz, the input windows are resampled using the Fourier method from a size of 4000 to a fixed size of 512, which allows for more efficient processing during model training, and it is the same input window size as the PPG-Dalia dataset.

3.3. Implementation

3.3.1. Representation learning

The hyperparameters and the architecture of the proposed deep learning model are determined by systematically searching through all possible combinations to obtain the best performance on the classification task using Leave-One-Subject-Out cross-validation (LOSO). Eventually, we used a CNN architecture deep learning model consisting of a 1D convolutional neural network layer with a series of five-layer blocks followed by a fully connected layer and a final classification layer. The layer blocks are composed of two 1D convolutional layers, each followed by the Exponential Linear Unit (ELU) activation function and, in the end, a MaxPooling layer. After the final layer block, there is a fully connected layer with a size of 64, which is the learned representation size, followed by the Rectified Linear Unit (ReLU) activation function. Finally, there is a classification layer (SoftMax activation function) with a size of 4, corresponding to the four categories. The final implemented CNN model details are available in Appendix.

The model is optimized using categorical cross-entropy as the loss function. The Adam optimizer is used with a learning rate of 0.00001 and a decay rate of 0.0001 for the PTT-PGG dataset and a learning rate of 0.0001 and a decay rate of 0.001 for the PPG-Dalia dataset. The batch size is 64, and training runs for 400 epochs for both datasets. To assess the randomness of the deep learning framework, each training process for each test subject is repeated five times. To evaluate the signal transformation classification performance, we use the Area under the ROC curve (AUC-ROC) metric.

3.3.2. Anomaly detection

In our PCA-based anomaly detection approach, we optimize the number of principal components by ensuring they cumulatively account for 99% of the data variance. The Isolation Forest model was implemented with 100 base estimators in the ensemble. The number of base estimators was chosen based on our preliminary experiments, which showed good performance in this setting. The Multivariate Normal Distribution-based anomaly detector was implemented utilizing a Gaussian Mixture Model with a single component. The parameters of this distribution, namely the mean vector and the covariance matrix, are learned directly from the data. In the evaluation phase, we assess the performance of our anomaly detectors by calculating the AUC-ROC.

3.4. Anomaly detection evaluation scenarios

We consider two evaluation scenarios for anomaly detection tasks: Generalization and Personalization.

3.4.1. Generalization scenario

In the generalization scenario, we aim to test the ability of the anomaly detection model to generalize across different individuals. For activity movement detection, shown in Fig. 2 (a), we train the model using data from all subjects performing *Sitting* activity as the main normal activity. This data is considered as 'normal training samples'. In the test phase, we introduce data from both a new activity, referred to as the 'anomalous activity' (in this case, *Walking*), and the main activity of a new subject (left-out) who was not part of the training data. We repeat this process for each subject, treating them as the test set (left-out subject), using the LOSO setting. We then calculate the mean and standard deviation of the performance metrics across all test sets.

For biometric identification, shown in Fig. 2 (b), we train the model on data from a group of subjects, who we refer to as the 'intended users'. This data forms our 'normal training samples'. We set aside 20% of the data from each subject for testing, using a 5-fold cross-validation approach. During the testing phase, we introduce 'anomalous data' from a new subject who is not part of the training data. This subject is referred to as the 'left-out' subject (user). To assess how well our model can differentiate the new user from the intended users, we use LOSO validation to treat each subject once as a 'left-out' user. We then calculate the mean and standard deviation of the performance metrics across all test sets.

3.4.2. Personalization scenario

In the personalization scenario, we aim to tailor the anomaly detection model to individual characteristics, both for movement detection and biometric identification. For activity movement detection, shown in Fig. 3 (a), we select one subject and train the model on data related to the main activity, which is *Sitting*. This data forms our 'normal training samples'. We reserve 20% of this data for testing, using a 5-fold cross-validation approach. During the testing phase, we introduce 'anomalous activity data' from the same subject, in this case, *Walking* activity, and we use 20% of this data using 5-fold cross-validation for testing. Thereby, our test set includes 20% of '*Walking* activity' data and 20% of the '*Sitting* activity' data from one selected subject. We repeat this process for each subject. Finally, we calculate the mean performance and standard deviation across all test sets.

For biometric identification, shown in Fig. 3 (b), we train the model using data from a single selected subject, who we refer to as the 'intended user'. This data forms our 'normal training samples'. We reserve 20% of this data for testing, using a 5-fold cross-validation method. During the testing phase, we introduce 'anomalous data' from a new subject who was not part of the training data, and we use 20% of its data using 5-fold cross-validation for testing. We compare each subject with the intended user in a pairwise manner. The average performance of these comparisons is taken as the performance of the intended user. We repeat this entire process for each individual, treating them as the intended user each time. Finally, we calculate the mean performance and standard deviation across all intended individuals.

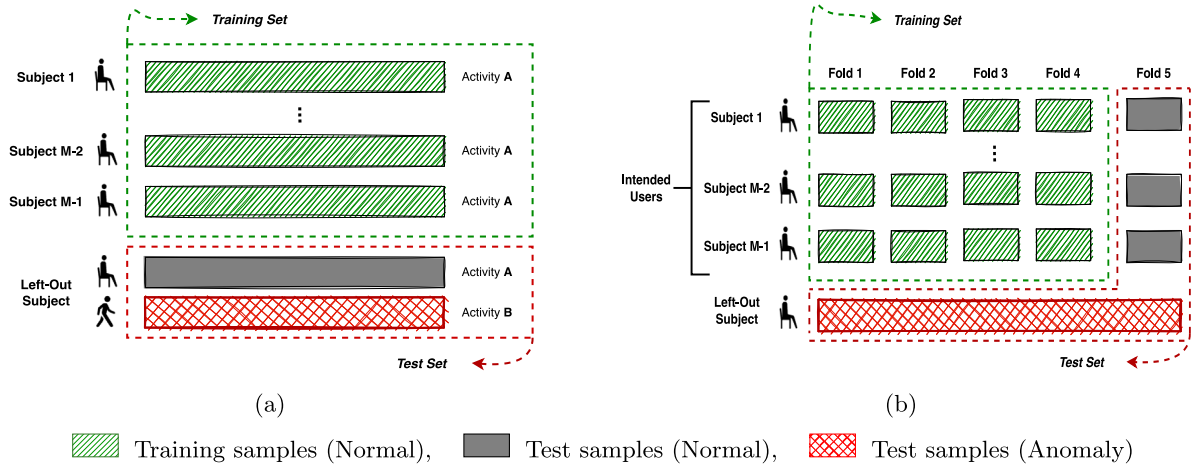

Fig. 2. *Overview of generalization scenario.* (a) Generalization in movement detection task (b) Generalization in biometric identification task. Note that the distribution of the anomalous and normal samples in training and test sets follows the same ratios as depicted in the figures.

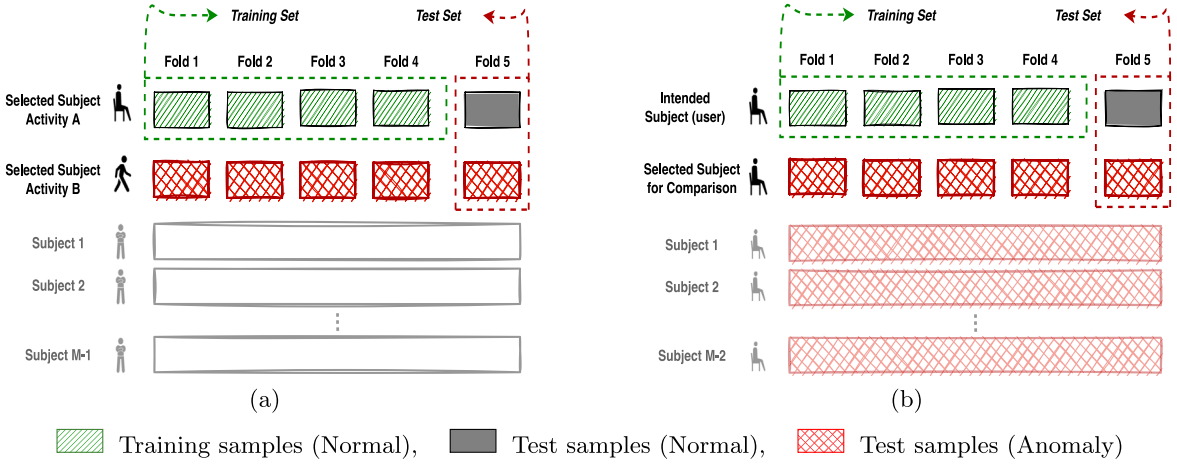

Fig. 3. *Overview of personalization scenario.* (a) Personalization in movement detection task. (b) Personalization in biometric identification task. Note that the distribution of the anomalous and normal samples in training and test sets follows the same ratios as depicted in the figures.

4. Results

4.1. Representation learning

The first step of the proposed framework is Representation Learning (see Fig. 1). The overall performance of the signal transformation classification task for both datasets is calculated across all test subjects. Both datasets have a high mean AUC of 0.92 ± 0.09 for the PTT-PPG and 0.93 ± 0.06 for the PPG-Dalia. These results indicate that the model is able to generalize to new data (subject) and accurately classifies the augmented and original PPG segments in both datasets.

4.2. Anomaly detection

4.2.1. Activity movement detection

The second step of the proposed framework is Anomaly Detection (see Fig. 1). Table 1 shows the results of movement detection for both datasets. In the generalized scenario, representation learning significantly improves the AUC performance for all three anomaly detection methods, suggesting its effectiveness in detecting anomalies in PPG data compared to the original data representation. For instance, the results of all anomaly detectors with PTT-PPG reveal that the AUC performance for anomaly detection barely reaches 0.5. However, the performance is increased towards 0.9 when using the learned representations.

Further, the original representation models demonstrate notable instability, indicative of high inter-subject variability. For example, the standard deviation in anomaly detectors without representation learning is around 0.4. However, employing representation learning substantially reduces this variability to below 0.1. These results underscore the capability of representation learning to facilitate better generalization across different individuals.

We also investigated the performance of the proposed methods in a personalized setting (Table 1). In the personalized setting, all methods show enhanced performance compared to the generalized scenario, with a moderate reduction in variability among individual performances. Notably, the integration of personalization with representation learning yields the most significant improvements. Here, the AUC performance approaches 1.0, and subject variability is markedly decreased to below 0.05. This demonstrates that the combined use of representation learning and personalization not only improves performance but also ensures consistency across individuals, effectively capturing subject-specific characteristics of PPG data.

4.2.2. Biometric identification

In light of the improved performance achieved through personalization in activity movement detection, biometric identification ensures that the detected anomalies are specific to the intended user. Table 2 shows the results of biometric identification in both generalized and personalized scenarios during *Sitting* Activity. In the generalized

Table 1
Mean test AUC-ROC performance of movement detection in the generalized and personalized scenario. The 'RL-' prefix designates anomaly detectors that employ learned representations from Representation Learning (RL) instead of the original data representation.

Anomaly Detectors	Movement detection AUC-ROC performance			
	PTT-PPG dataset		Dalia dataset	
	Generalized	Personalized	Generalized	Personalized
MVN	0.40 ± 0.39	0.74 ± 0.29	0.78 ± 0.16	0.93 ± 0.08
RL-MVN	**0.92 ± 0.09**	**0.98 ± 0.03**	**0.93 ± 0.10**	**0.97 ± 0.03**
IF	0.28 ± 0.34	0.37 ± 0.31	0.56 ± 0.14	0.87 ± 0.15
RL-IF	**0.91 ± 0.11**	**0.97 ± 0.05**	**0.88 ± 0.13**	**0.93 ± 0.03**
PCA	0.44 ± 0.37	0.81 ± 0.25	0.76 ± 0.19	0.94 ± 0.07
RL-PCA	**0.91 ± 0.09**	**0.97 ± 0.03**	**0.90 ± 0.18**	**0.95 ± 0.03**

Table 2
Mean test AUC-ROC performance of biometric identification in the generalized and personalized scenarios. Results are based on *Sitting* activity. The 'RL-' prefix designates anomaly detectors that employ learned representations from Representation Learning (RL) instead of the original data representation.

Anomaly Detectors	Biometric identification AUC-ROC performance			
	PTT-PPG dataset		Dalia dataset	
	Generalized	Personalized	Generalized	Personalized
MVN	0.40 ± 0.26	0.76 ± 0.22	0.43 ± 0.26	0.56 ± 0.20
RL-MVN	**0.60 ± 0.22**	**0.86 ± 0.08**	**0.55 ± 0.17**	**0.78 ± 0.09**
IF	0.45 ± 0.36	0.58 ± 0.29	0.45 ± 0.29	0.56 ± 0.24
RL-IF	**0.61 ± 0.24**	**0.86 ± 0.08**	**0.53 ± 0.18**	**0.74 ± 0.09**
PCA	0.39 ± 0.34	0.67 ± 0.24	0.43 ± 0.26	0.55 ± 0.22
RL-PCA	**0.59 ± 0.20**	**0.84 ± 0.09**	**0.55 ± 0.16**	**0.78 ± 0.09**

scenario, it can be observed that using representation learning is effective, and it improves the performance of all anomaly detectors across both datasets. While representation learning has been successful in improving performance, it still may not be perfect. This can be attributed to the inter-subject variability present in the data, as the model must distinguish between multiple people considered normal, which is challenging.

Considering the personalized scenario results, the performance of all methods is significantly higher compared to the generalized scenario, with substantially reduced variability: while in the generalized scenario, the standard deviations of the results are often around 0.2, in the personalized scenario, it is reduced to below 0.1. Moreover, representation learning continues to improve performance in the personalized setting, demonstrating the effectiveness of learned representations. These results indicate that minimizing inter-subject variability allows the model to better identify the anomalous person as it is easier to detect the anomalous individual from only one individual compared to a group.

4.3. Robustness of representation dimensionality

One key aspect of our anomaly detection framework is the dimensionality of the learned representation, denoted as $\mathbf{h}_i$. Fig. 4 illustrates the mean AUC-ROC performance of anomaly detectors with varying $\mathbf{h}_i$ dimensions ranging from 2 to 512 for both datasets in generalized and personalized scenarios.

As the $\mathbf{h}_i$ dimensionality increases, the AUC also increases up to a certain point. This trend suggests that as the dimensionality rises, the representation captures more valuable information for anomaly detection. However, once we reach a certain dimensionality, further increases do not provide additional benefits, and the performance is stable.

In both scenarios, the learned representation improves the AUC compared to the original signal. Results show that using learned representation leads to better performance when the dimensionality of $\mathbf{h}_i$ is reduced to extremely low levels. For instance, at the low dimensionality of 2 in PTT-PPG and 8 in the Dalia datasets, we can see the improvements in using learned representation over the original signal. Even when the dimensionality of the learned representation is the same as the original signal's dimensionality (512), it outperforms the original signal. This robust performance of the representation learning approach highlights its effectiveness in capturing the essential structure and patterns of the data and learning useful features across a wide range of low and high dimensionalities.

Choosing the right dimensionality depends on a balance between model performance and computational efficiency. Based on the results, the dimensionality of 64 for the PTT-PPG and 256 for the Dalia dataset seems to offer an ideal balance between computational efficiency and performance.

5. Discussion and conclusion

This paper proposes a framework for anomaly detection in PPG data, consisting of two stages: representation learning and anomaly detection (Activity Movement Detection and Biometric Identification). We tested the ability of the proposed framework in generalized and personalized scenarios. Our research demonstrates that through representation learning and person-specific models, we can effectively address the key challenges in analyzing PPG signals, such as inter-subject variability, avoiding the influence of factors like color and skin thickness, weight, bone structure, etc. This significantly enhances the accuracy of anomaly detection, potentially allowing for the detection of rare and subtle anomalous patterns.

The results from the activity movement detection highlight the effectiveness of representation learning in improving AUC performance and decreasing inter-subject variability. However, it is important to note the variations in the extent of variability reduction between datasets. For instance, while representation learning significantly decreases variability in the PTT-PPG dataset, the reduction in the Dalia dataset is less pronounced. This difference could be attributed to various factors, such as the inherent complexities of each dataset, differences in signal quality, sensor types, environmental conditions during data collection, and participant demographics. Despite these challenges, representation learning not only consistently improved performance across datasets but also effectively reduced intra-subject variability. However, for further improvements, personalization combined with representation learning emerged as a critical factor. By customizing models to individuals, we achieve more consistent results and effectively capture subject-specific characteristics. These findings underscore the potential for further advancements in personalization to address the challenges in variability, especially in complex datasets and in building reliable and accurate PPG-based health monitoring systems. Note that the original representation AUC results, which are lower than 0.5, may indicate that flipping the label is actually beneficial. For example, it can be seen in Table 1 that an AUC of 0.28 in the generalized scenario from the PTT-PPG dataset would significantly improve to 0.72 by flipping, but it still remains worse than the 0.91 obtained by the RL representation. These results suggest that our approach may be beneficial in real-world applications.

In addition to the findings from the Activity Movement Detection task, our results in Biometric Identification further emphasize the key role of representation learning and personalization. Similar to the previous task, representation learning significantly enhanced the performance in both generalized and personalized scenarios for biometric identification. Using personalization along with representation learning shows markedly higher performance and reduced variability. This improvement is attributed to the model's focus on individual characteristics. The consistency in these findings across both tasks underscores the efficacy of these methods separately and also in combination.

To further validate the effectiveness of our approach, we conducted a quantitative comparison with existing studies in PPG data for similar

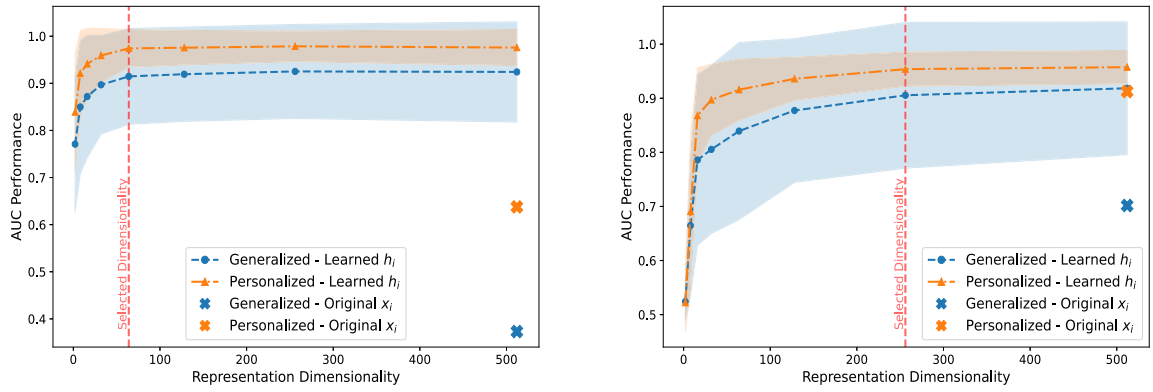

Fig. 4. *Overview of movement detection performance in the generalized and personalized scenarios for varying dimensions of the learned representation.* (a) the PTT-PPG dataset and (b) the Dalia dataset. The crosses indicate the performance obtained with the original representation.

tasks and scenarios. For instance, in movement detection, a study [31], employing fully supervised learning, reported an AUC of 0.89 in personalized and 0.78 in generalized scenarios on the Dalia dataset (although in a more complex multi-class classification setting). Our unsupervised approach with representation learning achieved AUCs of up to 0.93 in the generalized scenario and 0.97 in the personalized scenario, suggesting higher overall performance. Similarly, for biometric identification, a supervised study [32] reported an AUC of 0.72 ± 0.14 in personalized scenarios (on a different dataset). Our framework, however, achieved a higher AUC and lower inter-subject variability, with an AUC of 0.86 ± 0.08. Although these numbers cannot be fairly compared, the results underscore the potential of unsupervised learning methods in PPG anomaly detection, particularly when combined with representation learning and personalization.

Analyzing the robustness of representation dimensionality underscores its significance in anomaly detection frameworks. The performance of anomaly detection in relation to the dimensionality of the learned representation follows a pattern of initial gains followed by a plateau. This pattern suggests that while increasing dimensionality can enhance performance, there is a threshold beyond which additional increases do not yield further benefits. Interestingly, at the same dimensionality as the original signal, representation learning performs better. This suggests that learned representations can capture the (non-linear) underlying patterns or structures in the data that may not be immediately apparent in the original signal. Furthermore, the fact that the learned representation can outperform the original signal even at extremely low dimensionalities signifies that representation learning can effectively extract and retain the most critical information from the original signal, thereby enhancing anomaly detection.

In all experiments of our framework, the performance difference between anomaly detection methods was relatively small. Therefore, we cannot draw a clear conclusion about which method performs better than the others overall. It seems that the crucial point in anomaly detection is not the method, but it is the representation and the personalization.

Despite the promising results in using representation learning and personalization, it is important to note that further research is needed to evaluate the effectiveness of RL on a wider range of different types of real-world anomalies in PPG. This is particularly important for practical applications, such as using smartwatches and self-monitoring for anomaly detection in healthcare, where the complexity and variability of real-world anomalies may be high. Exploring different algorithms or techniques to enhance the learned representations of PPG data can also be a future direction to further improve anomaly detection performance and decrease inter-subject variability.

In conclusion, our proposed framework provides a promising approach for different types of anomaly detection in PPG data. Combination of representation learning and personalization provides a more effective approach for developing reliable, robust, and accurate health monitoring systems.

CRediT authorship contribution statement

Ramin Ghorbani: Methodology, Writing – original draft, Writing – review & editing, Visualization. **Marcel J.T. Reinders:** Supervision, Writing – review & editing. **David M.J. Tax:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets analyzed in the current study are publicly available and have been cited in the paper. The code used for analysis is available on the corresponding author's GitHub page.

Acknowledgment

Funding

This work was supported by the Dutch Research Council (NWO) [grant numbers 628.011.214].

Appendix. Deep learning model architecture for representation learning task

The Deep Learning framework is developed and evaluated in Python (Version 3.8.8) using Keras API. It should be noted that for the parameters that are not mentioned in the implementation details, the Keras default settings are used. The detailed architecture and outline of the implemented model for the representation learning task is shown in Table A.1 and Fig. A.1.

Table A.1
The detailed architecture of implemented encoder-classifier deep learning model..

Layer (Type)	Output shape	Parameters #
Encoder:		
Input layer	(None, 512, 1)	0
Conv1D layer (kernel-size = 64)	(None, 449, 32)	2080
Conv1D layer (kernel-size = 64, padding = "same")	(None, 449, 32)	65 568
Activation layer	(None, 449, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 449, 32)	65 568
Activation layer	(None, 449, 32)	0
Max Pooling1D layer	(None, 224, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 224, 32)	65 568
Activation layer	(None, 224, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 224, 32)	65 568
Activation layer	(None, 224, 32)	0
Max Pooling1D layer	(None, 112, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 112, 32)	65 568
Activation layer	(None, 112, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 112, 32)	65 568
Activation layer	(None, 112, 32)	0
Max Pooling1D layer	(None, 56, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 56, 32)	65 568
Activation layer	(None, 56, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 56, 32)	65 568
Activation layer	(None, 56, 32)	0
Max Pooling1D layer	(None, 28, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 28, 32)	65 568
Activation layer	(None, 28, 32)	0
Conv1D layer (kernel-size = 64, padding = "same")	(None, 28, 32)	65 568
Activation layer	(None, 28, 32)	0
Max Pooling1D layer	(None, 14, 32)	0
Flatten	(None, 448)	0
Fully connected layer	(None, 64)	28 736
Classifier:		
Fully connected output layer - (Activation = "softmax")	(None, 4)	260
Total Params: 686,756 Trainable Params: 686,756 & Non-Trainable Params: 0		

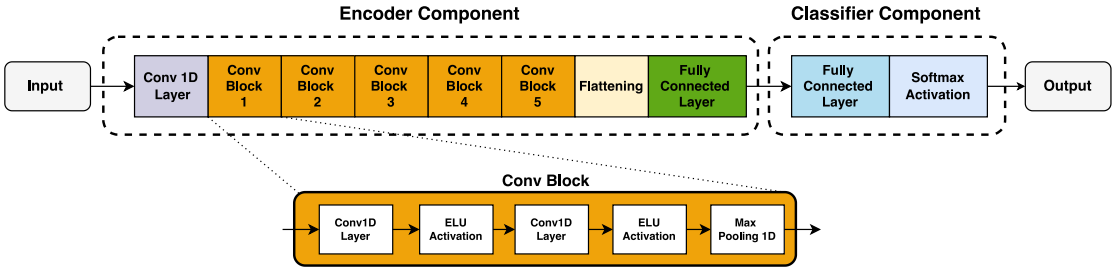

Fig. A.1. *Outline of the implemented encoder-classifier deep learning model.*

References

[1] John Allen, Photoplethysmography and its application in clinical physiological measurement, Physiol. Meas. 28 (3) (2007) R1.

[2] Ramin Ghorbani, Marcel J.T. Reinders, David M.J. Tax, Self-supervised ppg representation learning shows high inter-subject variability, in: Proceedings of the 2023 8th International Conference on Machine Learning Technologies, 2023, pp. 127–132.

[3] Malak Fora, Sajidah Al-Hammouri, Awad Al-Zaben, ECG abnormality detection from PPG signal, in: 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT, IEEE, 2019, pp. 103–106.

[4] Zeinab Rezaei Yousefi, Jakub Parak, Adrian Tarniceriu, Jarkko Harju, Arvi Yli-Hankala, Ilkka Korhonen, Antti Vehkaoja, Atrial fibrillation detection from wrist photoplethysmography data using artificial neural networks, in: World Congress on Medical Physics and Biomedical Engineering, Vol. 2019, 2018, pp. 399–404.

[5] Ming-Zher Poh, Yukkee Cheung Poh, Pak-Hei Chan, Chun-Ka Wong, Louise Pun, Wangie Wan-Chiu Leung, Yu-Fai Wong, Michelle Man-Ying Wong, Daniel Wai-Sing Chu, Chung-Wah Siu, Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms, Heart 104 (23) (2018) 1921–1928.

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

[7] Soonil Kwon, Joonki Hong, Eue-Keun Choi, Euijae Lee, David Earl Hostallero, Wan Ju Kang, Byunghwan Lee, Eui-Rim Jeong, Bon-Kwon Koo, Seil Oh, et al., Deep learning approaches to detect atrial fibrillation using photoplethysmographic signals: algorithms development study, JMIR mHealth uHealth 7 (6) (2019) e12770.

[8] Raghavendra Chalapathy, Sanjay Chawla, Deep learning for anomaly detection: A survey, 2019, arXiv preprint arXiv:1901.03407.

[9] Varun Chandola, Arindam Banerjee, Vipin Kumar, Anomaly detection: A survey, ACM Comput. Surv. (ISSN: 0360-0300) 41 (3) (2009) http://dx.doi.org/10.1145/1541880.1541882.

[10] Mohamed Elgendi, Richard Fletcher, Yongbo Liang, Newton Howard, Nigel H Lovell, Derek Abbott, Kenneth Lim, Rabab Ward, The use of photoplethysmography for assessing hypertension, NPJ Digit. Med. 2 (1) (2019) 60.

[11] Rado Kotorov, Lianhua Chi, Min Shen, et al., Personalized monitoring model for electrocardiogram signals: diagnostic accuracy study, JMIR Biomed. Eng. 5 (1) (2020) e24388.

[12] Yoshua Bengio, Aaron Courville, Pascal Vincent, Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.

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

[14] Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.

[15] Wenrui Zhang, Shijia Geng, Shenda Hong, A simple self-supervised ECG representation learning method via manipulated temporal–spatial reverse detection, Biomed. Signal Process. Control 79 (2023) 104194.

[16] Behzad Bozorgtabar, Dwarikanath Mahapatra, Guillaume Vray, Jean-Philippe Thiran, Salad: Self-supervised aggregation learning for anomaly detection on x-rays, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 468–478.

[17] Abinav Ravi Venkatakrishnan, Seong Tae Kim, Rami Eisawy, Franz Pfister, Nassir Navab, Self-supervised out-of-distribution detection in brain CT scans, 2020, arXiv preprint arXiv:2011.05428.

[18] He Zhao, Yuexiang Li, Nanjun He, Kai Ma, Leyuan Fang, Huiqi Li, Yefeng Zheng, Anomaly detection for medical images using self-supervised and translation-consistent features, IEEE Trans. Med. Imaging 40 (12) (2021) 3641–3651.

[19] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, Tomas Pfister, Cutpaste: Self-supervised learning for anomaly detection and localization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 9664–9674.

[20] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moeslund, Mubarak Shah, Self-supervised predictive convolutional attentive block for anomaly detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 13576–13586.

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

[22] Junjie Xu, Yaojia Zheng, Yifan Mao, Ruixuan Wang, Wei-Shi Zheng, Anomaly detection on electroencephalography with self-supervised learning, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2020, pp. 363–368.

[23] Yuxin Zhang, Jindong Wang, Yiqiang Chen, Han Yu, Tao Qin, Adaptive memory networks with self-supervised learning for unsupervised anomaly detection, IEEE Trans. Knowl. Data Eng. (2022).

[24] Pritam Sarkar, Ali Etemad, Self-supervised ECG representation learning for emotion recognition, IEEE Trans. Affect. Comput. 13 (3) (2020) 1541–1554.

[25] David Martinus Johannes Tax, One-class classification: Concept learning in the absence of counter-examples, 2002.

[26] Fei Tony Liu, Kai Ming Ting, Zhi-Hua Zhou, Isolation forest, in: 2008 Eighth IEEE International Conference on Data Mining, IEEE, 2008, pp. 413–422.

[27] Philip Mehrgardt, Matloob Khushi, Simon Poon, Anusha Withana, Pulse transit time PPG dataset, 2022.

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

[29] Zhilin Zhang, Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction, IEEE Trans. Biomed. Eng. 62 (8) (2015) 1902–1910.

[30] Seyed MA Salehizadeh, Duy Dao, Jeffrey Bolkhovsky, Chae Cho, Yitzhak Mendelson, Ki H Chon, A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intense physical activities using a wearable photoplethysmogram sensor, Sensors 16 (1) (2015) 10.

[31] Mahsa Sadat Afzali Arani, Diego Elias Costa, Emad Shihab, Human activity recognition: a comparative study to assess the contribution level of accelerometer, ECG, and PPG signals, Sensors 21 (21) (2021) 6997.

[32] Dwaipayan Biswas, Luke Everson, Muqing Liu, Madhuri Panwar, Bram-Ernst Verhoef, Shrishail Patki, Chris H Kim, Amit Acharyya, Chris Van Hoof, Mario Konijnenburg, et al., CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment, IEEE Trans. Biomed. Circuits Syst. 13 (2) (2019) 282–291.