Physiological
Measurement

PAPER • OPEN ACCESS

Generative adversarial networks with fully connected layers to denoise PPG signals

To cite this article: Itzel A Avila Castro *et al* 2025 *Physiol. Meas.* **46** 025008

IOP Publishing Physiol. Meas. 46 (2025) 025008 https://doi.org/10.1088/1361-6579/ada9c1

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
19 September 2024

REVISED
16 December 2024

ACCEPTED FOR PUBLICATION
13 January 2025

PUBLISHED
11 February 2025

PAPER

Generative adversarial networks with fully connected layers to denoise PPG signals

Itzel A Avila Castro[1,*], Helder P Oliveira[2,4], Ricardo Correia[1], Barrie Hayes-Gill[1], Stephen P Morgan[1], Serhiy Korposh[1], David Gomez[1] and Tania Pereira[3,4,*]

1 Optics and Photonics Group and Centre for Healthcare Technologies, University of Nottingham, Nottingham, United Kingdom
2 Faculty of Science, University of Porto and Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Porto, Portugal
3 Faculty of Science and Technology, University of Coimbra, Coimbra, Portugal
4 Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Porto, Portugal
* Authors to whom any correspondence should be addressed.

E-mail: itzel.castro@nottingham.ac.uk and tania.pereira@inesctec.pt

Keywords: generative adversarial networks, denoising signals, corrupted PPG signal, movement artifacts

Supplementary material for this article is available online

Abstract

*Objective.*The detection of arterial pulsating signals at the skin periphery with Photoplethysmography (PPG) are easily distorted by motion artifacts. This work explores the alternatives to the aid of PPG reconstruction with movement sensors (accelerometer and/or gyroscope) which to date have demonstrated the best pulsating signal reconstruction. *Approach.* A generative adversarial network with fully connected layers is proposed for the reconstruction of distorted PPG signals. Artificial corruption was performed to the clean selected signals from the BIDMC Heart Rate dataset, processed from the larger MIMIC II waveform database to create the training, validation and testing sets. *Main results.* The heart rate (HR) of this dataset was further extracted to evaluate the performance of the model obtaining a mean absolute error of 1.31 bpm comparing the HR of the target and reconstructed PPG signals with HR between 70 and 115 bpm. *Significance.* The model architecture is effective at reconstructing noisy PPG signals regardless the length and amplitude of the corruption introduced. The performance over a range of HR (70–115 bpm), indicates a promising approach for real-time PPG signal reconstruction without the aid of acceleration or angular velocity inputs.

1. Introduction

Nowadays photoplethysmography (PPG) is a versatile wearable technology to monitor a variety of physiological conditions such as blood oxygen saturation, heart rate (HR), blood pressure, cardiac output, arterial stiffness and respiration (Allen 2007, Almarshad *et al* 2022, Charlton *et al* 2023, Ferizoli *et al* 2024). This technology uses two optoelectronic components: a light source and a photodetector, where the light sources illuminates the skin, and the transmitted or reflected light from the skin tissue is converted into an electrical signal by the photodetector, to photocurrent linked with the blood volume changes. The waveform of the signal detected by the photodetector is a peripheral pulse synchronized to each heartbeat. This waveform comprises pulsating (AC) and non-pulsating (DC) components (Webster 1997, Caizzone *et al* 2019). The morphology of the pulsating components due to arteries shares the same frequency as the HR, and the non-pulsating component can be due to the light reflected by static tissues such as bone, venous blood due to physiological processes such as respiration, vasomotor activity and thermal regulation. Furthermore, this DC component rises with movement between the sensor and the skin surface, where the increase in the magnitude of this movement can distort the PPG signal, affecting the accuracy of the physiological parameters extrapolated from this waveform such as the HR (Pereira *et al* 2020).

Numerous techniques have been implemented to tackle motion artifacts. These techniques can involve either software signal processing tools and/or hardware conditioning. Rudimentary software approaches extend from the removal of the corrupted segments through manual identification and signal quality assessment through classification methods (Asgari *et al* 2009, Sukor *et al* 2011, Pereira *et al* 2019). More robust methods have also been developed to reconstruct the corrupted segments. Combining independent component analysis and block interleaving with low pass filtering the motion artifact was reduced exploiting the independence in the periodicity of the PPG signal (Kim and Yoo 2006); iterative motion artifact removal using the singular spectral analysis algorithm for an accurate estimation of HR and SpO_2 values (Salehizadeh *et al* 2014); TROIKA (signal decompositTion for denoising, sparse signal RecOnstruction for hIgh resolution for spectrum estimation, and spectral peaK tracking with verification) (Zhang *et al* 2015); robust preprocessing through wavelet denoising for HR estimation (Mullan *et al* 2015). PPG signal reconstruction can be achieved through empirical mode decomposition (EMD) and discrete wavelet transform (DWT) (Tang *et al* 2017), obtaining the relevant component of PPG signals when the noise degree is low. Motion signals from accelerometers and gyroscopes have been used as a reference along with adaptive filtering for motion artifact removal (Ribeiro *et al* 2023); auto-regressive model and Kalman filters have been used for signal reconstruction correlating the signal using the preceding information, and filtering the sudden movements at the same time (Nooralishahi *et al* 2019). Deep learning Techniques have also been implemented to reconstruct PPG signals such as Autoencoders (Lee *et al* 2019, Jain *et al* 2024).

Moreover, generative models have become fundamental in the field of imaging, demonstrating remarkable capabilities in improving image resolution, enhancing visual quality, and addressing specific challenges such as restoring blurred regions, upscaling low-resolution images, and normalizing staining colours (Seo *et al* 2021). For example, enhancing super-resolution performance and computational efficiency, as seen in PCA-SRGAN (Dou *et al* 2020), which employs principal component analysis to reduce the dimensionality of feature maps extracted from low-resolution images. Recent advancements in generative models have further expanded their potential. For instance, cross-domain translation techniques, as exemplified by CycleGAN (Dou *et al* 2020), enable image-to-image translation without the need for paired datasets, thereby broadening their applicability across diverse imaging domains. Furthermore, the integration of generative adversarial networks (GANs) with other generative frameworks, such as GPT and diffusion models, has extended their utility to areas like natural language processing and audio generation (Bengesi *et al* 2024). In the biomedical domain, novel implementations like Asymmetric CycleGAN networks have been utilized for tasks such as smoke removal in endoscopic images (Zhou *et al* 2024). Similarly, in photoplethysmogram (PPG) signal reconstruction, Cycle-GANs have been employed to convert PPG signals into 2D representations for reconstruction using unpaired datasets (Zargari *et al* 2021). Additionally, Deep Convolutional GANs (DC-GANs) have been trained on clean PPG signals to generate reconstructed signals by averaging the predicted and raw data, resulting in improved reconstruction accuracy (Wang *et al* 2022).

However, the best performance overall in PPG signal reconstruction is still achieved with the aid of an accelerometer and gyroscope (Zhang *et al* 2015). Whereas, the signal reconstruction using DC-GANs still need to average the signal reconstructed with the raw data to improve its reconstruction. Considering these drawbacks, there is still a need for solutions to receive a corrupted signal and recover the noisy parts without using any additional parts of the signal or any other dependency from other signals such as ECG. This kind of approach will be very useful for a real-time application since the read and recovery of the signal can be done in a semi-continuous way, using a sliding window that the recovery system will process. Furthermore, this approach would be beneficial for the cases where PPG measurements occur at remote locations inside the body where with limited space the use of an accelerometer and/or gyroscope would be impossible at the same location. This frequently happens when using optical fibre probes for physiological measurement inside the trachea and oesophagus (Kyriacou 2005, Correia *et al* 2022).

The objective of this current study is to develop a GAN model trained with pairs of clean and noisy signals and learn to recover a corrupted signal into a clean one. The methodology proposed is fully detailed in the materials and methods section. It involves the selection of the PPG database, signal quality assessment of this database to extract all the clean segments, followed by artificial corruption to obtain pairs of corrupted and clean segments which are the inputs for training, validation and testing of the proposed network which is a GAN with fully connected layers (FC-GAN). These results are evaluated in section 3 with Bland–Altman analyses of the extracted HR along with different metrics: Euclidean Distance, signal to noise ratio (SNR) and mean absolute error (MAE) comparing with ECG and PPG ground truth HR.

2. Materials and methods

For the development of the learning model to reconstruct the corrupted signals, the pipeline presented in figure 1 was developed.

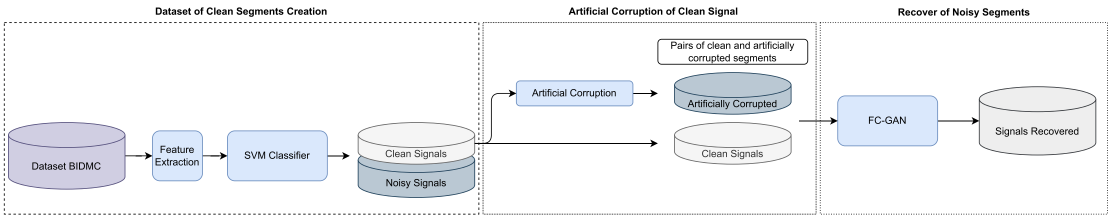

Figure 1. Pipeline of the methodology to develop a generative adversarial network with fully connected layers (FC-GAN) model to denoise the corrupted PPG signals. This work was composed of three main steps: the development of a dataset of clean signals; and the artificial corruption of those clean signals to create a pair of clean-corrupted signals to train the FC-GAN to learn how to denoise the signals.

2.1. PPG database

The dataset used in this work was the BIDMC Heart Rate Dataset (Pimentel *et al* 2017), which is derived from the larger MIMIC II waveform database. BIDMC Heart Rate Dataset comprises 7949-time series data segments of 32 s window length from 53 patients. This dataset constitutes PPG signals sampled at 125 Hz and HR values from ECG signals. The BIDMC dataset was segmented into 8 s segments, resulting in a dataset with 31 796 segments.

2.2. Signal quality assessment

A signal quality assessment is performed to the BIDMC Heart Rate Dataset to extract the clean PPG signals. The rationale behind the selection of clean signals is to create a dataset where the noise can be controlled to generate a pairwise input to the proposed network, a corrupted PPG segment (input to be reconstructed) along with the corresponding clean PPG segment (ground-truth/target). This dataset with noise control relies on using clean segments to perform a controlled noise corruption where the target/clean signal is known. To obtain a set with only clean segments a support vector machine (SVM) classification model was implemented. This classifier uses five input features extracted from the PPG signals: the variation in Kurtosis, the variation in Skewness, the variation in the approximate entropy of the cardiac cycles, the Shannon entropy and the Spectral entropy (Mahmoudzadeh *et al* 2021). To train and test the SVM classifier, a small set of 300 segments from BIDMC Heart Rate dataset with noisy and clean signal labels was generated with manual labelling, identifying the clean and noisy PPG segments with small amplitude DC variations (DC amplitude of maximum 16% to maximum amplitude overall of the whole segment) and not combined with other higher-frequency signals. After this labelling criteria, the five features previously mentioned were extracted for each segment. The training parameters of SVM classifier were a regularization parameter $C = 0.1$ and random state $= 0$ and the training and test sets were defined following 75:25 ratio. This trained SVM classifier model was further applied to the extracted features of the rest of the dataset (31 496 segments after excluding the number of segments used train and test the SVM classifier). After applying the classification to the remaining 31 496 segments, clean PPG segments with a confidence score greater than 1.85 were selected as clean PPG segments, resulting in a clean data set of 11 718 PPG segments.

The confidence of the classified clean segments was increased by only selecting the clean segments with a confidence score above 1.85. To calculate this score, a decision function was applied to the classifier giving a score to each segment classified. This score is calculated as the distance that separates the features of each classified segment from the hyperplane. The larger values mean the features of the segment are farther from this hyperplane of the support vector model and is classified with higher confidence. The calculated scores varied from 0.03 to 2.25, where 1.85 demonstrated the score limit for a classification that best approximate the criteria defined for clean PPG segments.

2.3. Artificial corruption of clean data set

An Artificial Corruption algorithm was implemented on the clean PPG signals previously identified by the SVM classifier from the BIDMC dataset, to create a new dataset composed of clean and corrupted segments to be used for training, validation, and testing of the FC-GAN model. The corruption of the clean signals was done at three different lengths to the last 2, 4, and 6 s. This algorithm first consisted in the random generation of a sine wave with a normal distribution of frequencies between 0.01–10 Hz, frequencies usually observed in PPG motion artifacts (Bagha and Shaw 2011, Rojano and Isaza 2016, Lee *et al* 2020) and four different weight amplitudes (1, 0.5, 0.33, and 0.25). Finally, this randomly generated sine wave was added to the last 250 (2 s), 500 (4 s), and 750 (6 s) data points of the clean segments, resulting in 3 types of corrupted signals. The SNR of every PPG corrupted signal was determined. It is important to note that the noise from the randomly generated sine wave corresponds to the range of frequencies 0.01–10 Hz, which is generally

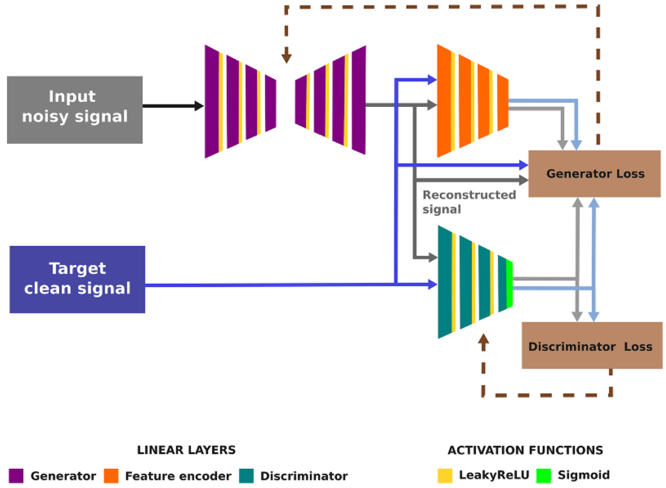

Figure 2. Generative adversarial network with fully connected layers (FC-GAN). The Generator uses an auto encoder structure (encoder–decoder, each one of four layers, with LeakyReLU activation on the first three layers for both structures), and an additional feature encoder structure is used to compare additional features between our target and reconstructed signal that aids the optimisation of the generator Loss. Then the Discriminator consists of an encoder of four layers with Leaky ReLU layers in the first three layers and a Sigmoid function in the fourth layer that aids the discrimination of the reconstructed signals by comparing the target signal with the generated/reconstructed signal.

related to motion artifacts. However, this range of frequencies covers also the frequency of the baseline wander noise (<0.1 Hz), the frequencies of the respiration-induced variations (0.1–0.5 Hz) and the frequencies of the instrumentation noise (>5 Hz) (Awodeyi *et al* 2014).

2.4. FC-GAN approach

The FC-GAN model used in this work was inspired by Akcay *et al* (2018), Wang *et al* (2022), where they implement the GANomaly architecture[5] which consists in an Autoencoder as the Generator, followed by a Feature Encoder to extract further information and an Encoder as the Discriminator to optimize the reconstructed segment. This work is novel as it replaces the convolutional layers with linear layers (figure 2). It is assumed that linear layers may offer advantages processing one-dimensional signals, with no spatial hierarchy as they can easily capture global patterns and dependencies, treating all parts of the signal with equal importance, connecting every neuron in one layer to every neuron in the next layer (Schwing and Urtasun 2015, Vo and Lee 2018), in contrast to Convolutional Layers that prioritize local information which it is useful in higher dimensional data. Single Layer Perceptron model, the basic structure of the linear layer in fully connected layers has demonstrated its robustness to exploit the global information of one dimensional data, utilising a diffuse reflectance spectrum to predict scattering and absorption coefficients (Fernandes *et al* 2021).

The Generator follows an encoder–decoder structure where the encoder and decoder comprise both four linear layers where in both cases the first three layers follow a LeakyReLU activation function. The feature encoder has the same structure as the encoder of the Generator. The encoder of the discriminator consists of four linear layers, where the first three include each a LeakyReLU activation function, while the last layer follows a sigmoid function. All the LeakyReLU activation functions were set with a negative slope of 0.5 with the operation in-place.

2.5. Loss functions

Two main loss functions were implemented Generator L_G and Discriminator loss L_D, similar to the GANomaly approach implemented by Akcay *et al* (2018). The Discriminator loss was minimised at the first iteration between batches, while the Generator loss was minimised in the remaining iterations in the training period. The Generator loss is a weighted linear combination of an encoder loss L_{enc}, contextual loss L_{con}, and adversarial loss L_{adv}, for a batch of size N,

$$L_G = w_{enc} L_{enc} + w_{con} L_{con} + w_{adv} L_{adv}. \tag{1}$$

The encoder loss uses the outputs of the feature encoder (x, encoded features of the reconstructed signal and y, encoded features of the target signal) into a Smooth L1 loss function. This function combines the

[5] This work is available on https://github.com/openvinotoolkit/anomalib.

advantages of both L1 loss (MAE) for large errors $|x_n - y_n| \geqslant \beta$ and L2 loss mean squared error (MSE) for small errors $|x_n - y_n| < \beta$, using β as a regularisation parameter,

$$L_{\text{enc}} = \ell(x,y) = \{l_1, \ldots, l_N\}^{\text{T}} \tag{2}$$
$$l_n = \begin{cases} 0.5\,(x_n - y_n)^2/\beta, & \text{if } |x_n - y_n| < \beta \\ |x_n - y_n| - 0.5 \times \beta, & \text{otherwise.} \end{cases}$$

The contextual loss takes the output of the reconstructed x and the target signal y to compute the negative value of the SNR,

$$L_{\text{SNR}}\text{con} = -\text{SNR}(x,y) = -10\log_{10}\left(\frac{P_y}{P_{\text{noise}}}\right) \tag{3}$$
$$\text{where, } P_{\text{noise}} = \text{noise}^2, \;\; P_y = y^2, \;\; \text{noise} = y - x.$$

The adversarial loss used the outputs of the Discriminator and computes a MSE loss.

$$L_{\text{adv}} = \ell(x,y) = \{l_1, \ldots, l_N\}^{\text{T}}, \tag{4}$$
$$l_n = (x_n - y_n)^2.$$

Meanwhile, the Discriminator loss uses the output of the Discriminator layers to compute a binary cross-entropy loss between the target and reconstructed input probabilities,

$$L_{\text{D}} = \ell(x,y) = \{l_1, \ldots, l_N\}^{\text{T}}, \tag{5}$$
$$l_n = -y_n \cdot \log x_n + (1 - y_n) \cdot \log(1 - x_n).$$

2.5.1. Implementation details

The Smooth L1, MSE and binary cross-entropy losses were implemented using the Pytorch libraries of each function. The β value for Smooth L1 loss function was set by default 1.0.

The Discriminator loss uses as an input the binary output of the Discriminator model which judge if the reconstructed segments look like real or fake. Meanwhile, the Generator loss uses as an input the encoded features of the reconstructed segment. During the training stage, the model alternates between training the Discriminator loss (first iteration) and the Generator loss (subsequent iterations). This balances the learning of both optimizations to prevent one from overpowering the other.

2.6. Training and application

To train the FC-GAN model to denoise the PPG artificial corrupted segments, the input pairs: artificial corrupted PPG segments, and the correspondent target clean PPG segments were normalized in a range from 0–1 using z-score standardization. The distribution of data among the training, validation, and test sets followed an 80:10:10 ratio. The training set consisted of 9376 inputs. The training process ran 1000 epochs in total with a batch size of 356. Adam Optimizer from Pytorch library (Paszke *et al* 2019) was used as an optimization algorithm, with the Generator and Discriminator learning rates were set at 1×10^{-4}, $\beta_1 = 0.5$ and $\beta_2 = 0.999$. These parameter values showed the best convergence rates among different combinations tested.

The validation process was performed on 1172 input pairs, 10% of the dataset. The trained parameters were saved for the testing process and the remaining 10% of the dataset was tested. For every reconstructed segment after introducing the corrupted segments to the model in the testing phase, the SNR was determined.

3. Results and discussion

3.1. Performance metrics

The performance of the methodology proposed was analysed through different approaches. Figure 3 shows four different examples of PPG signal reconstruction, comprising examples that show the capacity to recover from noisy signals (figures 3(a) and (b)) being able to extract similar HR as the fiducial points of the PPG pulses matches. The degradation of the capacity of reconstruction in figure 3(b) can be attributed to the corruption of the signal at a lower frequency within the frequency of the HR. On the other hand, figures 3(c) and (d) depicted the worst scenario where the reconstruction of the PPG signal struggled to match the fiducial points of the clean PPG signal. It is noted that in this scenario the target HR were 63.5 bpm and 123

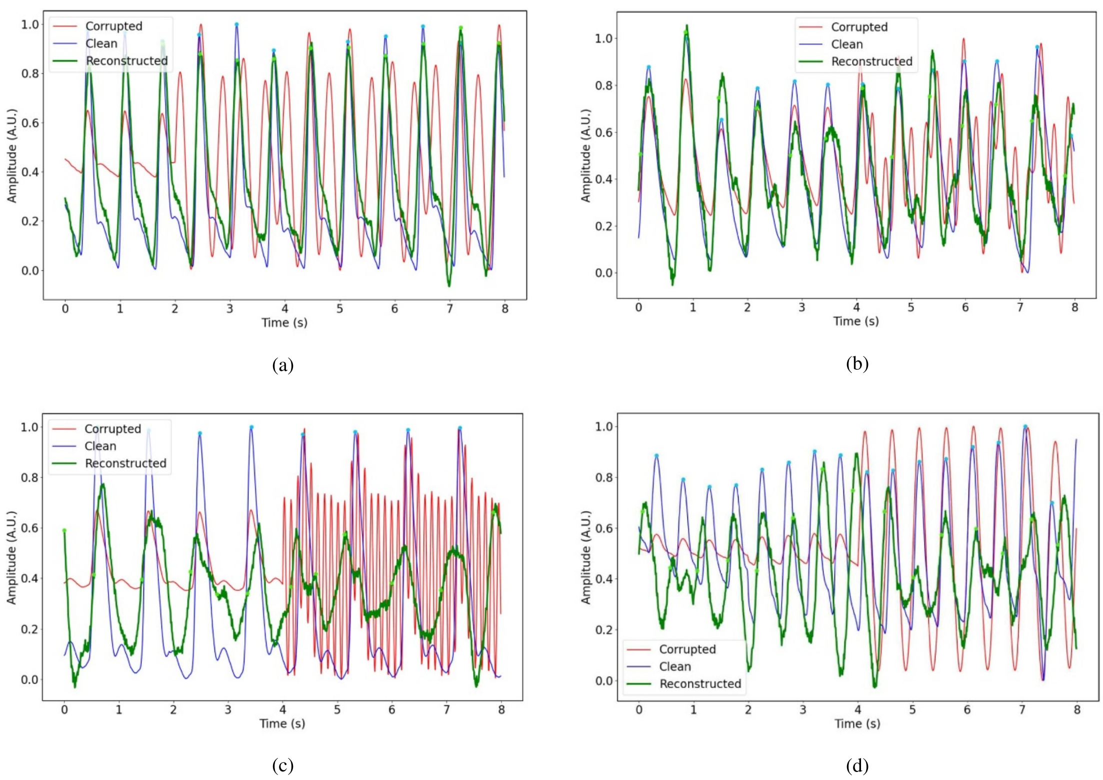

Figure 3. Example of the original signal (clean) marked as blue, the corrupted marked as red, and the reconstructed one marked as green, for the different values of HR. (a) PPG signals corrupted in last 6 s. The recovered HR was 88.24 bpm for a target HR 88.24 bpm, the SNR improved from −8.44 (noisy signal) to 9.95 (recovered signal), the frequency of the noisy signal is around 2 times higher than the frequency of the target PPG signal. (b) PPG signals corrupted in last 4 s. The recovered HR was 94.95 bpm for a target HR 93.17 bpm, the SNR improved from −1.87 (noisy signal) to 12.02 (recovered signal), the frequency of the noise signal is around 3 times higher than the frequency of the target PPG signal. (c) PPG signals corrupted in last 4 s. The recovered HR was 75.76 bpm for a target HR 63.56 bpm, the SNR improved from −6.3 (noisy signal) to 3.31 (recovered signal), the frequency of the noise signal is around 10 times higher than the frequency of the target PPG signal. (d) PPG signals corrupted in last 4 s. The recovered HR was 114.51 bpm for a target HR 123 bpm, the SNR improved from −9.37 (noisy signal) to 5.56 (recovered signal), the frequency of the noise signal matches the frequency of the target PPG signal.

bpm, representing cases at the limits of human HR taking into consideration the population of HR used for the training phase.

Two Bland–Altman analyses were performed to compare the measurement of HR from the PPG signal (original and reconstructed) and ECG. The first Bland–Altman plot (figure 4(a)) compares the HR from the reconstructed PPG signals and ECG HR from the BIDMC Heart Rate Dataset. The second Bland–Altman plot (figure 4(b)) compares the HR from the reconstructed PPG signals and the HR from the target PPG signals. Furthermore, a correlation analysis (figure 4(c)) between the reconstructed PPG signals and the HR from the target PPG signals was computed. From figures 4(a) and (b), it was noticed a large difference (greater than ±32 bpm) between reconstructed and target HR under certain circumstances when the target PPG HR was outside the range between 70 and 115 bpm as it is observed in figure 4(c).

Figure 5 shows the HR distribution of the training and testing set, where the distribution of the ECG HR of the training set differs from the distribution of the ECG HR of the testing set, which explains why the developed model lacks performance outside the range between 70 and 115 bpm since it did not have enough data to be trained under these scenarios.

The results from the test set were evaluated by computing the Euclidean Distance between the PPG reconstructed signal and the PPG target signal. The SNR of the reconstructed signal was also computed. Moreover, two different MAE were calculated. The first was by comparing the HR of the PPG reconstructed signal and the ECG HR from the BIDMC Heart Rate Dataset, and the second MAE value was by comparing the HR of the PPG reconstructed signal and the HR of the target PPG signal. After noticing the pattern of the reconstruction results were divided into two sets. Set 1 corresponds to the population of signals from the testing set within a HR between 70 and 115 bpm, where the model showed a significant in reconstruction (average Euclidean Distance between clean and reconstructed PPG signal of 3.85 and MAE of the reconstructed PPG HR against the clean PPG HR of 1.31 bpm). Set 2 corresponds to the population of signals from the testing set with a HR outside the range between 70 and 115 bpm, in this case the model

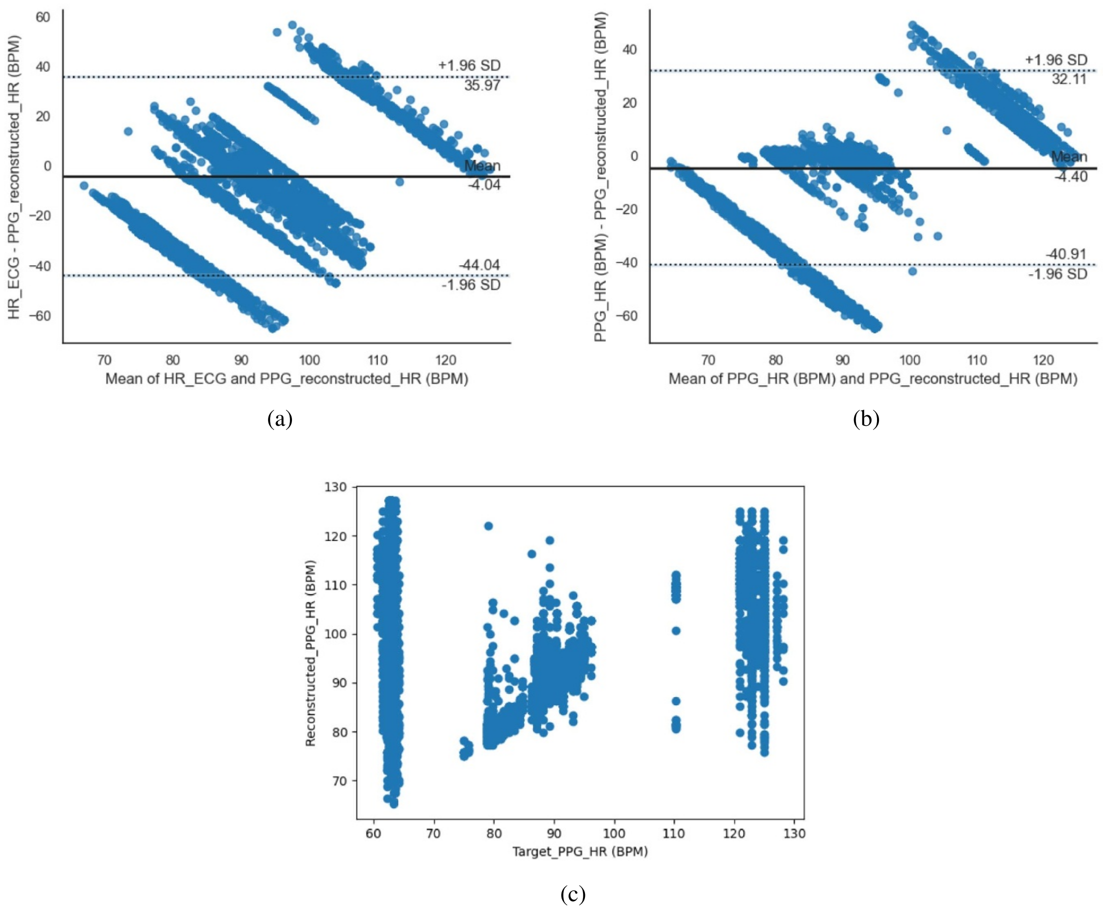

Figure 4. Comparing HR of PPG and ECG signals. (a) Bland Altman analysis between ECG and reconstructed PPG HR. (b) Bland Altman analysis between target and reconstructed PPG HR. (c) Correlation analysis between ECG and PPG reconstructed HR. The Bland Altman plot shows a great amount of data (69% in (a) and 67% in (b) from the total of the test set) inside the limits of ± 1.96 SD. The values outside of the limit follow a distribution that is not verified in the training dataset, and probably, for this reason, the reconstructed signal was not so well reconstructed. SD: Standard Deviation.

showed a poor reconstruction (MAE of the reconstructed PPG HR against the clean PPG HR of 29.1 bpm and a Euclidean Distance between signals of 9.5) due to a lack of generalization in the training set. Table 1, shows the results for these different sets considering all the signals with different lengths of corruption (2, 4, 6 s) at different amplitudes of corruption (weight = 1, 0.5, 0.33, and 0.25). Table 1 synthesises the results at all different combinations where the mean Euclidean Distance was 3.85 ± 1.33 for Set 1 and 9.5 ± 0.97 for Set 2. The mean SNR was 11.8 ± 2.4 for Set 1 and 3.55 ± 0.95 for Set 2, the MAE against ECG HR is 12.8 bpm for Set 1 and 20.8 bpm for Set 2 and the MAE against target PPG HR is 1.31 bpm for Set1 and 29.1 bpm for Set 2. The metrics were also computed for all different cases (lengths and weight of corruption) where there is no significant variation between these cases. Table 2 shows an example of the metrics computed for all different corrupted lengths (2–6 s), where the amplitude of the corruption was assigned a weight = 1.

3.2. Comparison with previous works

Table 3 describes the state of the art of the outcomes in PPG signal reconstruction. These articles were selected since they share the same objective as this work proposes, besides the metrics employed by the authors are comparable with the implemented in this work.

Kim and Yoo (2006) combined independent component analysis (ICA) and a signal enhancement method to separate the PPG signals from the motion artifacts. They obtained a Mean Squared Error (MSE) from 0.4–4 between the reference and reconstructed PPG signals. The MSE increased as the power of the motion artifacts and the amount of frequency overlapping the clean signal and the motion artifacts increased.

Salehizadeh *et al* (2014) used an iterative motion artifact removal (IMAR) approach, using singular spectral analysis to reduce motion and noise artifacts. They found a statistical significance $p > 0.05$ in the HR of the reconstructed signals in 7 out of 9 subjects.

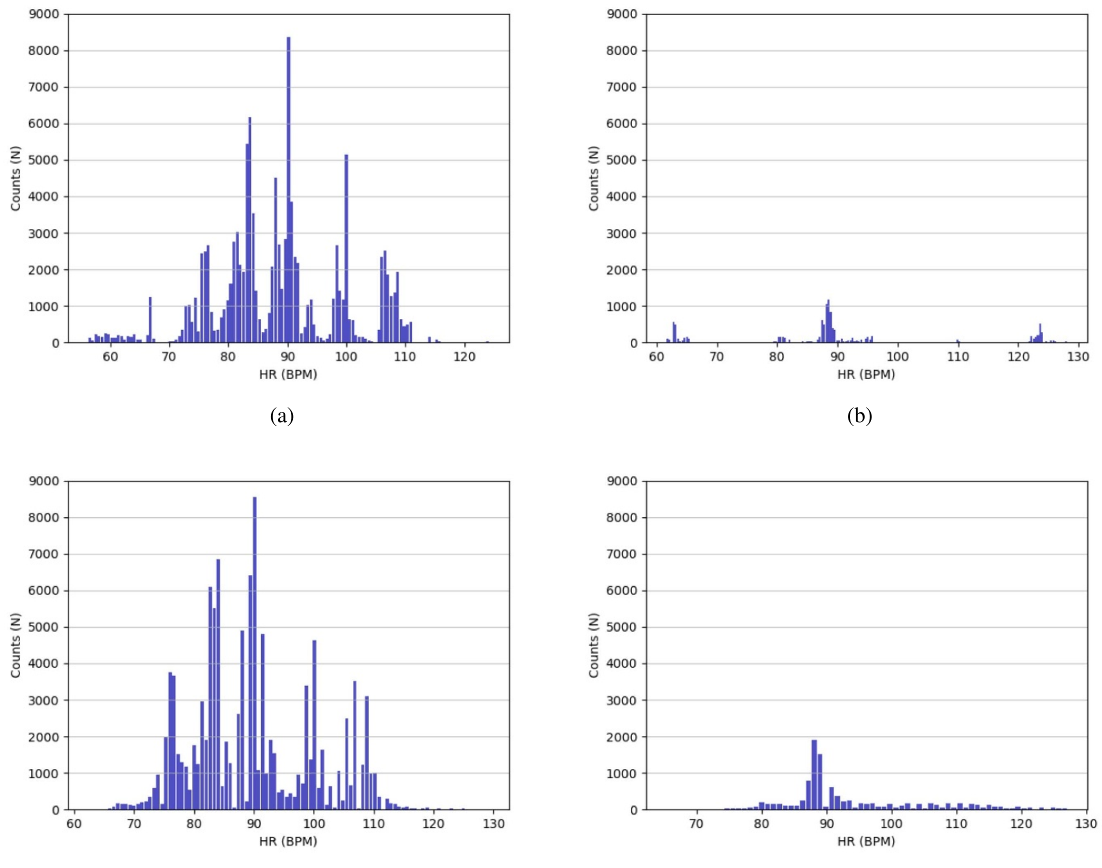

Figure 5. Histograms of HR in training and testing data sets from ECG and Reconstructed PPG signals. (a) ECG HR histogram in BIDMC Heart Rate training set. (b) ECG HR histogram in BIDMC Heart Rate testing set. (c) Reconstructed PPG HR histogram in the training set.(d) Reconstructed PPG HR histogram in the testing set.

Table 1. Assessment of the quality of the recovery signal based on different metrics. Set 1 corresponds to the population of signals from the testing set within a HR between 70 and 115 bpm and Set 2 corresponds to the population of signals from the testing set with a HR outside the range between 70 and 115 bpm.

Metric	Set 1	Set 2
Euclidean Distance	3.85 ± 1.33	9.5 ± 0.97
SNR(dB)	11.80 ± 2.4	3.55 ± 0.95
MAE (against ECG HR)	12.8 bpm	20.8 bpm
MAE (against PPG HR)	1.31 bpm	29.1 bpm

Table 2. Dataset with the last 2, 4, and 6 s corrupted, weight = 1.

	2 s		4 s		6 s	
Metric	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2
Euclidean Distance	4.22 ± 1.58	9.71 ± 1.08	4.27 ± 1.94	9.64 ± 1.24	3.88 ± 1.17	9.22 ± 0.63
SNR(dB)	11.05 ± 2.67	3.37 ± 1.19	11.08 ± 2.9	3.45 ± 1.25	11.67 ± 2.37	3.88 ± 0.99
MAE against ECG HR (bpm)	12.76	20	12.9	19.68	12.83	20.7
MAE against PPG HR (bpm)	1.43	27.65	1.66	28.86	1.35	25.9

Mullan *et al* (2015) used a wavelet-based denoising method with the aid of acceleration data, they evaluated the performance of the HR and reached a MAE 1.96 ± 2.86 bpm and Pearson correlation coefficient $r = 0.98$.

Zhang *et al* (2015) combined ensemble empirical model decomposition with spectrum subtraction technique along with acceleration signals. They calculated an average absolute error (Aerror) across 12 subjects 1.83 ± 1.21 bpm, average percentage error of 1.40% and Pearson correlation coefficient of $r = 0.989$.

Table 3. Results comparison of the state-of-the-art techniques for denoising PPG signals with motion artifacts. Different metrics were employed by these authors for performance evaluation, mean squared error (MSE) by Kim and Yoo between the reference and denoised PPG signal, mean absolute error (MAE) by extracting the HR of the reconstructed PPG signals. Average absolute error (Aerror) and Average percentage error, Bland–Atlman, and Correlation analysis with the HR extracted from the reconstructed PPG signals. Mean Sum Error (MSumE) by Tang between the reference HR and the HR of the reconstructed PPG signal. peak-peak error (PPE) and root mean squared error (RMSE) between the generated and reference signals.

Paper	Method	Outcome
Kim and Yoo (2006)	ICA	MSE 0.4–4
Salehizadeh *et al* (2014)	IMAR	HR with $p > 0.05$ in 7 out of 9 subjects
Mullan *et al* (2015)	Wavelet and acceleration	MAE 1.96 ± 2.86 bpm $r = 0.98$
Zhang *et al* (2015)	EMD and acceleration	Aerror: 1.83 ± 1.21 bpm Average error%: 1.4 LOA: -7.56, 6.61 bpm $\sigma = 3.62$ $r = 0.989$
Tang *et al* (2017)	DWT & EMD	MSumE 2.95–32.24 bpm
Lee *et al* (2019)	BRDAE	7.9 dB SNR improved
Zargari *et al* (2021)	CycleGAN	PPE: 0.95 bpm RMSE: 2.18 bpm
Wang *et al* (2022)	DC-GAN	PPE:.7–1.9 s

Tang *et al* (2017) combined DWT and EMD to reconstruct corrupted PPG signals and evaluated its performance with the Mean Sum Error (MSumE) of the HR estimated, the MSumE varied from 2.95 bpm to 32.94 bpm. Lee *et al* (2019) used a Bidirectional recurrent auto-encoder (BRDAE) obtaining an average SNR improvement of 7.9 dB in the validation set. Zargari *et al* (2021) used a Cycle Generative Adversarial Network and obtained a peak-to-peak error (PPE) of 0.95 bpm and RMSE of 2.18 bpm. Wang *et al* (2022) proposed DC-GAN, where the PPE varied from 0.7 to 1.9 s.

Current results in the trained scenario (Set 1) are comparable to these results obtained in the previous works in particular to Mullan *et al* (2015) and Zhang *et al* (2015) where they obtained a MAE of 1.96 bpm and Aerror of 1.83 bpm while in our case is 1.31 bpm. The main advantage of our approach is that we are not aiding the reconstruction with an accelerometer signal to gain an effective reconstruction. However, the downside of the developed methodology is the generalization since the uneven HR distribution of our dataset lacks signal outside 70 and 115 bpm.

Furthermore, the DC-GAN (Wang *et al* 2022) and Cycle GAN (Zargari *et al* 2021) networks were implemented to test our designed corrupted datasets for a fair comparison with our proposed model (FC-GAN). For the implementation of the DC-GAN and Cycle GAN models, these networks were adapted to our designed dataset by modifying the dimension of the input and output layers (both considering an input and output of 8 s) to match the length of the input pairs of our dataset. Additionally, in the Cycle GAN approach after the signal to image transformation, the image was resized from 1000 to 128 pixels to reduce the computational demand. Also, it was noticed that this model had better performance with a smaller training set of 870 segments than with training set size used for DC-GAN and FC-GAN models (9376 segments).

Table 4 shows the different metrics (Euclidean Distance, SNR and MAE against ECG and PPG HR) used to compare the performance of the different network models with our designed datasets with different lengths of corruption (2, 4, 6 s) at different amplitudes of corruption (weight = 1, 0.5, 0.33, and 0.25). Our proposed network (FC-GAN) demonstrates the lowest Euclidean Distance 3.85 ± 1.33 between the target and reconstructed signals, while DC-GAN reconstruction shows the largest Euclidean Distance with big error (13.4 ± 17.08). When comparing the SNR of the reconstructed signals the best SNR 11.80 ± 2.4 is shown in our proposed methodology (FC-GAN). In this case the SNR from DC-GAN and Cycle GAN have a difference of 1 dB where the Cycle GAN SNR shows less standard deviation. Evaluating the MAE against ECG HR the best performance was achieved by the Cycle GAN network with 9.65 bpm. Finally, when comparing the MAE against ECG HR the best performance was achieved by the DC-GAN model with 1.13 bpm, however with a difference of 0.20 bpm when compared with our proposed methodology.

3.3. Model's runtime

The real-time application of the FC-GAN model is evaluated by evaluating the execution runtime to reconstruct the signals of the designed datasets with different lengths of corruption (2, 4, 6 s) at different amplitudes of corruption (weight = 1, 0.5, 0.33, and 0.25). This runtime test was run on a computer with an Apple M2 Pro processor with 32 GB of RAM (only using the CPU). Overall, the mean (±std) time to reconstruct 8 s of signal at 125 Hz to the FC-GAN model proposed takes 18.637 ± 0.265 ms. Moreover, when comparing this reconstruction time on the same set of signals with others' networks from the literature (DC-GAN (Wang *et al* 2022) and Cycle-GAN (Zargari *et al* 2021)). DC-GAN's runtime is 24.486 ± 0.643 ms and Cycle-GAN's runtime is 1845.557 ± 103.597 ms. The runtime of the FC-GAN model demonstrates to be

Table 4. Results comparison of the implementation of Cycle GAN (Zargari *et al* 2021), DC- GAN (Wang *et al* 2022) and FC-GAN (proposed methodology) models using the same metrics in our designed corrupted datasets, Set 1 with HR between 70 and 115 bpm. Overall the best performance is shown in FC-GAN model considering all the metrics. However, the best MAE against ECG HR is shown by the Cycle GAN model 9.65 bpm and the best MAE against PPG HR is shown by the DC-GAN model 1.13 bpm with a difference 0.20 bpm with our proposed method FC-GAN with a MAE against PPG HR 1.31 bpm.

Metric	CycleGAN	DC-GAN	FC-GAN (proposed model)
Euclidean Distance	5.45 ± 0.94	13.4 ± 17.08	3.85 ± 1.33
SNR(dB)	9.12 ± 1.29	8.12 ± 3.13	11.80 ± 2.4
MAE (against ECG HR)	9.65 bpm	12.58 bpm	12.8 bpm
MAE (against PPG HR)	2.48 bpm	1.13 bpm	1.31 bpm

the fastest model among the other GAN-based models tested, the architecture of the FC-GAN model is the less complex among these models as it uses only linear layers and supports the ability to reconstruct the signal in real time since the delay to reconstruct the signal is approximately 429 times smaller than the size of the reconstruction.

3.4. Limitations and future work

The current work presents limitations, that have an impact on the results achieved. The dataset available to train the model was limited by a range of HR, between 60 to 120 bpm. The limits of the HR used for training the model generate a learning model that presents difficulties to recover signals outside of this range. This is a very common issue of deep learning models and is characterized by the lack of generalization for distributions out of the distribution. The distribution represented in figure 5 shows a different distribution of the test set compared with the training set used to train the learning model. Deep learning has achieved great performance based on independent and identically distributed (IID) assumptions, but the application of DL is more challenging for out-of-distribution (OOD) scenarios (Wu *et al* 2022, Zhang *et al* 2023).

The dataset used in the current work may not cover all the heterogeneities that can exist in the population, associated with all pathological cases and especially related to cardiovascular diseases that are responsible for the main changes in the PPG signal, such as the arrhythmias.

The combination of the recovery model with a noise selector can be helpful. A previous selection of the signals that need to be recovered will decrease the number of segments processed and can make the system of acquisition and processing more efficient for real-time analysis.

4. Conclusions

A FC-GAN was proposed for the first time to reconstruct PPG signals with motion and noise artifacts. Linear layers may offer advantages processing one-dimensional signals, as they can easily capture global patterns and dependencies, giving equal importance to the whole signal, in contrast Convolutional Layers tend prioritize local information with more relevance in higher dimensional data. Clean signals from BIDMC Heart Rate Dataset were extracted and artificially corrupted to train, validate, and test the proposed network. The developed approach showed great capability to reconstruct the PPG signal with a MAE of 1.31 bpm, particularly between HR of 70 and 115 bpm, based on the distribution of the ECG HR of the training set. Outside this range, the performance of the reconstruction is significantly reduced by a ratio of 22. Tackling the generalisation of the trained dataset might overcome this issue. Moreover, the FC-GAN along with the training framework, using a noise artificial controlled dataset demonstrated to be a promising approach to reconstruct corrupted PPG signals in real time, where it demonstrated the capability to reconstruct 8 s of signals in 18.637 ± 0.265 ms. The significant performance of this method is comparable with the state of art of techniques that reached a significant performance with the aid of acceleration signals.

Additionally, for a fair comparison with the other state of the art of GAN techniques, the designed dataset was used for signal reconstruction in DC-GAN and Cycle GAN models. The Cycle GAN implemented model shown insights to be a robust network for signal reconstruction, however this robustness was not fully exploited, the segments of the inputs from the dataset to the model had to be resized from 1000 to 128 pixels, as it requires high computational resources for training and application, making it at the same time less suitable for real-time signal reconstructions compared with DC-GAN and FC-GAN models. The implemented DC-GAN shows the lowest MAE against PPG HR (1.13 bpm). However, this same performance was not seen with the Euclidean Distance between the reconstructed and target signals and SNR of its reconstructed signals where the proposed FC-GAN model outperformed and its MAE differs by 0.20 bpm compared with DC-GAN.

Data availability statement

The data used in this study is publicly available through the PhysioNet repository. The BIDMC PPG and Respiration Dataset can be accessed at https://physionet.org/content/bidmc/1.0.0/. Researchers can freely download the data by agreeing to the PhysioNet data usage terms.

Acknowledgments

IAA-C thanks INESC TEC for the amazing opportunity to conduct this internship and a further collaboration that accomplished this research. To my colleagues, Luis Emanuel Pereira and Hugo Oliveira for the helpful conversations, and to my supervisors for their constant and valuable support and knowledge during this process. To the University of Nottingham and CONAHCYT (Mexico) for their financial support in pursuing my PhD. This work was supported by the Medical Research Council (UK) under Grant MR/T025638/1.

ORCID iDs

Itzel A Avila Castro https://orcid.org/0000-0001-5548-7264
Helder P Oliveira https://orcid.org/0000-0002-6193-8540
Ricardo Correia https://orcid.org/0000-0001-6405-5340
Barrie Hayes-Gill https://orcid.org/0000-0003-4237-7001
Stephen P Morgan https://orcid.org/0000-0003-4069-3801
Serhiy Korposh https://orcid.org/0000-0001-5179-6991
David Gomez https://orcid.org/0000-0003-1305-8474
Tania Pereira https://orcid.org/0000-0003-1681-2436

References

Akcay S, Atapour-Abarghouei A and Breckon T P 2018 GANomaly: semi-supervised anomaly detection via adversarial training (arXiv:1805.06725)

Allen J 2007 Photoplethysmography and its application in clinical physiological measurement *Physiol. Meas.* **28** R1

Almarshad M A, Islam M S, Al-Ahmadi S and BaHammam A S 2022 Diagnostic features and potential applications of PPG signal in healthcare: a systematic review *Healthcare* **10** 547

Asgari S, Bergsneider M and Hu X 2009 A robust approach toward recognizing valid arterial-blood-pressure pulses *IEEE Trans. Inf. Technol. Biomed.* **14** 166–72

Awodeyi A E, Alty S R and Ghavami M 2014 On the filtering of photoplethysmography signals *2014 IEEE Int. Conf. on Bioinformatics and Bioengineering* (IEEE) pp 175–8

Bagha S and Shaw L 2011 A real time analysis of PPG signal for measurement of SpO_2 and pulse rate *Int. J. Comput. Appl.* **36** 45–50

Bengesi S, El-Sayed H, Sarker M K, Houkpati Y, Irungu J and Oladunni T 2024 Advancements in generative AI: a comprehensive review of GANs, GPT, autoencoders, diffusion model and transformers *IEEE Access* **12** 69812–37

Caizzone A, Boukhayma A and Enz C 2019 AC/DC ratio enhancement in photoplethysmography using a pinned photodiode *IEEE Electron Device Lett.* **40** 1828–31

Charlton P H *et al* 2023 The 2023 wearable photoplethysmography roadmap *Physiol. Meas.* **44** 111001

Correia R *et al* 2022 Intra-tracheal multiplexed sensing of contact pressure and perfusion *Biomed. Opt. Express* **13** 48–64

Dou H, Chen C, Hu X, Jia L and Peng S 2020 Asymmetric CycleGAN for image-to-image translations with uneven complexities *Neurocomputing* **415** 114–22

Dou H, Chen C, Hu X, Xuan Z, Hu Z and Peng S 2020 PCA-SRGAN: incremental orthogonal projection discrimination for face super-resolution *Proc. 28th ACM Int. Conf. on Multimedia* pp 1891–9

Ferizoli R, Karimpour P, May J M and Kyriacou P A 2024 Arterial stiffness assessment using PPG feature extraction and significance testing in an *in vitro* cardiovascular system *Sci. Rep.* **14** 2024

Fernandes L, Carvalho S, Carneiro I, Henrique R, Tuchin V V, Oliveira H P and Oliveira L M 2021 Diffuse reflectance and machine learning techniques to differentiate colorectal cancer *ex vivo Chaos* **31** 053118

Jain P, Ding C, Rudin C and Hu X 2024 A self-supervised algorithm for denoising photoplethysmography signals for heart rate estimation from wearables *Harv. Data Sci. Rev.* **6**

Kim B S and Yoo S K 2006 Motion artifact reduction in photoplethysmography using independent component analysis *IEEE Trans. Biomed. Eng.* **53** 566–8

Kyriacou P A 2005 Pulse oximetry in the oesophagus *Physiol. Meas.* **27** R1

Lee J, Kim M, Park H K and Kim I Y 2020 Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths *Sensors* **20** 1493

Lee J, Sun S, Yang S M, Sohn J J, Park J, Lee S and Kim H C 2019 Bidirectional recurrent auto-encoder for photoplethysmogram denoising *IEEE J. Biomed. Health Inform.* **23** 2375–85

Mahmoudzadeh A, Azimi I, Rahmani A M and Liljeberg P 2021 Lightweight photoplethysmography quality assessment for real-time IoT-based health monitoring using unsupervised anomaly detection *Proc. Comput. Sci.* **184** 140–7

Mullan P *et al* 2015 Unobtrusive heart rate estimation during physical exercise using photoplethysmographic and acceleration data *2015 37th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* pp 6114–7

Nooralishahi P, Loo C K and Shiung L W 2019 Robust remote heart rate estimation from multiple asynchronous noisy channels using autoregressive model with Kalman filter *Biomed. Signal Process. Control* **47** 366–79

Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library *Advances in Neural Information Processing Systems* vol 32 (Curran Associates, Inc.) (available at: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf), pp 8024–35

Pereira T, Gadhoumi K, Ma M, Liu X, Xiao R, Colorado R A, Keenan K J, Meisel K and Hu X 2019 A supervised approach to robust photoplethysmography quality assessment *IEEE J. Biomed. Health Inform.* **24** 649–57

Pereira T, Tran N, Gadhoumi K, Pelter M M, Do D H, Lee R J, Colorado R, Meisel K and Hu X 2020 Photoplethysmography based atrial fibrillation detection: a review *npj Digit. Med.* **3** 3

Pimentel M A F, Johnson A E W, Charlton P H, Birrenkott D, Watkinson P J, Tarassenko L and Clifton D A 2017 Toward a robust estimation of respiratory rate from pulse oximeters *IEEE Trans. Biomed. Eng.* **64** 1914–23

Ribeiro L, Oliveira H, Hu X and Pereira T 2023 AI-based models to predict the heart rate using PPG and accelerometer signals during physical exercise *2023 IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)* pp 4398–403

Rojano J F and Isaza C V 2016 Singular value decomposition of the time-frequency distribution of PPG signals for motion artifact reduction *Int. J. Signal Process. Syst.* **4** 475–82

Salehizadeh S M, Dao D K, Chong J W, McManus D, Darling C, Mendelson Y and Chon K H 2014 Photoplethysmograph signal reconstruction based on a novel motion artifact detection-reduction approach. Part II: motion and noise artifact removal *Ann. Biomed. Eng.* **42** 11

Schwing A G and Urtasun R 2015 Fully connected deep structured networks (arXiv:1503.02351)

Seo J-W, Jung H-G and Lee S-W 2021 Self-augmentation: generalizing deep networks to unseen classes for few-shot learning *Neural Netw.* **138** 140–9

Sukor J A, Redmond S and Lovell N 2011 Signal quality measures for pulse oximetry through waveform morphology analysis *Physiol. Meas.* **32** 369

Tang S D, Goh Y S, Wong M D and Lew Y E L 2017 PPG signal reconstruction using a combination of discrete wavelet transform and empirical mode decomposition *2016 6th Int. Conf. on Intelligent and Advanced Systems (ICIAS)* (IEEE)

Vo D M and Lee S-W 2018 Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions *Multimedia Tools Appl.* **77** 18689–707

Wang Y, Azimi I, Kazemi K, Rahmani A M and Liljeberg P 2022 PPG signal reconstruction using deep convolutional generative adversarial network *2022 44th Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)* vol 2022-July (Institute of Electrical and Electronics Engineers Inc.) pp 3387–91

Webster J G 1997 *Design of Pulse Oximeters* (CRC Press)

Wu J, Zlobec I, Lafarge M, He Y and Koelzer V H 2022 Towards IID representation learning and its application on biomedical data (arXiv:2203.00332)

Zargari A H A, Aqajari S A H, Khodabandeh H, Rahmani A M and Kurdahi F 2021 An accurate non-accelerometer-based PPG motion artifact removal technique using CycleGAN (arXiv:2106.11512)

Zhang M, Yuan J, He Y, Li W, Chen Z and Kuang K 2023 MAP: towards balanced generalization of IID and OOD through model-agnostic adapters *Proc. IEEE/CVF Int. Conf. on Computer Vision* pp 11921–31

Zhang Y, Liu B and Zhang Z 2015 Combining ensemble empirical mode decomposition with spectrum subtraction technique for heart rate monitoring using wrist-type photoplethysmography *Biomed. Signal Process. Control* **21** 119–25

Zhou Y, Hu Z, Xuan Z, Wang Y and Hu X 2024 Synchronizing detection and removal of smoke in endoscopic images with cyclic consistency adversarial nets *IEEE/ACM Trans. Comput. Biol. Bioinform.* **21** 670–80