Biomedical Signal Processing and Control 85 (2023) 104992

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Rapid extraction of respiratory waveforms from photoplethysmography: A deep corr-encoder approach

Harry J. Davies *, Danilo P. Mandic

Electrical and Electronic Engineering, Imperial College London, United Kingdom

ARTICLE INFO

Keywords:
Photoplethysmography
Signal processing
Deep learning
Machine learning
Biomedical engineering
BIosignals
Wearable health
Wearables

ABSTRACT

Much of the information related to breathing is contained within the photoplethysmography (PPG) signal, through changes in venous blood flow, heart rate and stroke volume. We aim to leverage this fact, by employing a novel deep learning framework which is a based on a repurposed convolutional autoencoder. Our corr-encoder model aims to encode all of the relevant respiratory information contained within photoplethysmography waveform, and decode it into a waveform that is similar to a gold standard respiratory reference — the Capnogram. The model is employed on two photoplethysmography data sets, namely Capnobase and BIDMC. We show that the model is capable of producing respiratory waveforms that approach the gold standard, while in turn producing state of the art respiratory rate estimates. We also show that when it comes to capturing more advanced respiratory waveform characteristics such as duty cycle, our model is for the most part unsuccessful. A suggested reason for this, in light of a previous study on in-ear PPG, is that the respiratory variations in finger-PPG are far weaker compared with other recording locations. Importantly, our model can perform these waveform estimates in a fraction of a millisecond, giving it the capacity to produce over 6 hours of respiratory waveforms in a single second. Moreover, we attempt to interpret the behaviour of the kernel weights within the model, showing that in part our model intuitively selects different breathing frequencies. The model proposed in this work could help to improve the usefulness of consumer PPG-based wearables for medical applications, where detailed respiratory information is required.

1. Introduction

Wearable health technology promises to revolutionise medicine through avenues of personalised care and long term monitoring of chronic diseases. However, what wearables gain from convenience and quantity of data, they often lose from an available information perspective. A prime example of this is the use photoplethysmography (PPG) for respiratory applications. Whilst the use of PPG has so far demonstrated high accuracy in the monitoring of respiratory rate, much of the respiratory information encoded within PPG is still untapped. Increases in resting respiratory rate can be indicative of breathing disorders, such as chronic obstructive pulmonary disease (COPD), restrictive lung diseases such as pulmonary fibrosis and pulmonary hypertension [1]. Importantly, further breathing pattern information such as inspiratory duty cycle can help to distinguish between specific types of lung disease [2]. To this end, we focus on closing the gap between photoplethysmography and gold standard respiratory monitoring through a multi-layer convolutional encoder–decoder framework that takes PPG as an input and outputs highly accurate respiratory waveforms. We demonstrate that these respiratory waveforms are not just useful for respiratory rate estimation, but are also comparable in terms of mean absolute error with gold standard respiratory measures. Furthermore, this simple deep learning model is computationally cheap to run, making implementation in wearables highly feasible.

1.1. Photoplethysmography and respiration

Photoplethysmography (PPG) refers to a non-invasive measurement of blood volume. It operates by emitting a light through the skin and into the underlying tissue, with either one or multiple light emitting diodes, and then measuring the light either transmitted through the tissue (transmittance PPG) or reflected back (reflectance PPG) to one or multiple photodiodes. Blood absorbs light, and therefore when more blood is present less light is reflected back to the photodiode. Thus, PPG can measure changes in blood volume that occur with the pulse and respiration. There is a pressure gradient between the peripheral veins and the heart which results in the return of deoxygenated blood to the heart (known as venous return). Given that the heart lies within the thoracic cavity, and intrathoracic pressure must change to allow for the

* Corresponding author.
E-mail addresses: harry.davies14@imperial.ac.uk (H.J. Davies), d.mandic@imperial.ac.uk (D.P. Mandic).

https://doi.org/10.1016/j.bspc.2023.104992
Received 1 January 2023; Received in revised form 27 March 2023; Accepted 30 April 2023
Available online 13 May 2023

flow of air in and out of the lungs, respiration therefore modulates this pressure gradient and venous return. When we breathe out, intrathoracic pressure increases to force air out of the lungs and this in turn reduces the pressure gradient between the heart and peripheral veins. A smaller pressure gradient means less venous return to the heart, and an increase in blood volume at the site of the PPG sensor. Moreover, as the lungs constrict more blood flows from the pulmonary veins into the left atrium, then resulting in an increased stroke volume and thus an increased pulse amplitude at the site of the PPG sensor. A larger pulse amplitude with expiration is also accompanied by a slower pulse rate. All three of these changes are reversed during inspiration [3].

To recover respiratory information from PPG, techniques usually rely on the extraction of one or multiple of the three respiratory modes (intensity variations, pulse amplitude variations and pulse interval variations). Common methods of estimating respiratory rate include finding the spectral peak corresponding to respiration in the PPG signal [4] and band-pass filtering to detect respiratory peaks in the time domain [5]. Photoplethysmography can also be decomposed into physically meaningful frequency components (intrinsic mode functions) through the use of empirical mode decomposition (EMD) or its multivariate equivalent (MEMD) [6]. This has been combined with methods such as principal component analysis to select the respiratory modes to give a more accurate respiratory rate estimate [7]. It has also been shown that respiratory intrinsic mode functions are detailed enough to ascertain the ratio between the duration of inspiration and expiration, allowing for the classification of chronic obstructive pulmonary disease (COPD) [8]. The methods above are commonly applied to all of the three respiratory variations, and the resulting respiratory rate estimations can then be fused into a final more accurate respiratory rate estimation [9].

1.2. Capnography

Capnography refers to the measurement of CO_2 concentrations over time during breathing [10]. It is primarily measured from patients who are intubated during scenarios such as surgery. When a patient breathes in, the concentration of CO_2 measured is negligible, whereas when a patient breathes out the concentration of CO_2 is far higher as a result of metabolic processes. In general, this results in a waveform similar to a square wave that acts as a gold standard measure of breathing.

There are certain cases where capnography is atypical, such as a drop in the amplitude of end tidal CO_2 (possibly indicative of a heart failure or pulmonary embolism [11]). Another common example is where the typical square wave pattern becomes a "shark fin", and this indicates severe airway obstruction such as is the case with COPD. For the purposes of this work we consider how much of the typical capnography respiratory waveform is encoded within the photoplethysmogram, and not emergency medicine scenarios with changes in capnogram amplitude.

1.3. Autoencoders for biosignals

Typically, autoencoders are employed as an unsupervised machine learning technique to learn an efficient coding of an input. The autoencoder operates by compressing an input into a latent space, which should represent information that is fundamental to the input signal. This latent space is then upsampled back into an output, which in typical applications should be the same as the input. These efficient codings of the input that exist in the latent space can then be used directly for classification [12], or the output itself can be used for classification [13]. A typical extension of this principle is for denoising, whereby the input is corrupted by noise and the autoencoder is tasked to match the output with a non-corrupted version of the input. Examples include the denoising of medical images [14] and the denoising of biosignals such as the electrocardiogram (ECG) [15].

Fundamentally, the autoencoder structure works by learning a compressed principal relationship between the input and the output. When it comes to PPG, there is clearly respiratory information encoded within the signal. We therefore hypothesise that by using the encoder–decoder structure with photoplethysmography as an input, and training the output against a gold standard respiratory waveform in the form of capnography, it is likely that we can condense and extract most of the relevant respiratory information that is available in PPG. Given that this re-purposing of the conventional autoencoder now aims to extract correlated information between an input and reference signal, we refer to this model as a corr-encoder.

2. Methods

2.1. Datasets

For the initial leave-one-subject-out cross-validation of our model, we use the Capnobase "respiratory benchmark" dataset [16]. This data consists of 42 subjects (29 children and 13 adults) recorded for 8 minutes each during either spontaneous or controlled breathing. The data contains simultaneously recorded finger-based photoplethysmography and capnography, where the capnography serves as a gold standard respiratory reference. The Capnobase encompasses a broad range of respiratory rates, from below 10 breaths per minute to above 40 breaths per minute. The abnormal abundance of higher respiratory frequencies is likely due to a majority of pediatric patients, as children tend to have higher resting respiratory rates. Photoplethysmography can lose respiratory information at higher respiration rates given the low-pass transfer function between breathing and PPG [17]. Therefore, having an abundance of higher frequency respiration data for testing, such as is the case in the Capnobase data set, is crucial to ensure the robustness of an algorithm. Capnobase also contains numerous examples of artefacts where either the capnography or PPG signal is corrupted. It is common for studies to remove epochs of Capnobase recordings that contain artefacts [18,19], but in this study we choose to leave these recordings in our analysis for a more realistic view of performance.

For further validation of the model we test on the BIDMC PPG and respiration data set [20,21], which contains signals from the MIMIC-II data set. Performance of respiratory rate estimation algorithms on the BIDMC data set is generally worse than on the Capnobase data set, given the strong presence of arterial blood pressure variations at approximately 0.12 Hz which are difficult to distinguish from respiratory variations [20]. Respiration in the BIDMC data set is derived from two of the electrocardiogram leads being repurposed for impedance pneumography, which tracks the movement of the chest during respiration. Given that this is different in structure and delay to capnography, we cannot directly compare our predicted respiratory waveforms that are trained on capnography to the BIDMC respiratory waveforms in terms of mean absolute error. However, we can compare both waveforms visually and examine the frequency content for respiratory rate estimation. The primary goal of the model presented in this work is to extract the shared information between the respiratory reference and photoplethysmography which, independent of the type of reference signal, is still the respiratory information. The rationale behind both testing and retraining the same model on the BIDMC dataset was to test the ability of the model to generalise to new data having had its parameters fine tuned on Capnobase. As with the Capnobase data set, we use all of the data in our analysis and do not exclude artefacts. The BIDMC data is resampled to 30 Hz, to match the sampling rate of the downsampled Capnobase data.

The raw PPG from both the Capnobase and BIDMC data sets was high-pass filtered to remove any DC offset, with a cut-off frequency controlled by the pulse oximeter manufacturer [16]. As a further step, each window of PPG was re-scaled to between −1 and 1, and each window or capnography and impedance pneumography was re-scaled to between 0 and 1.

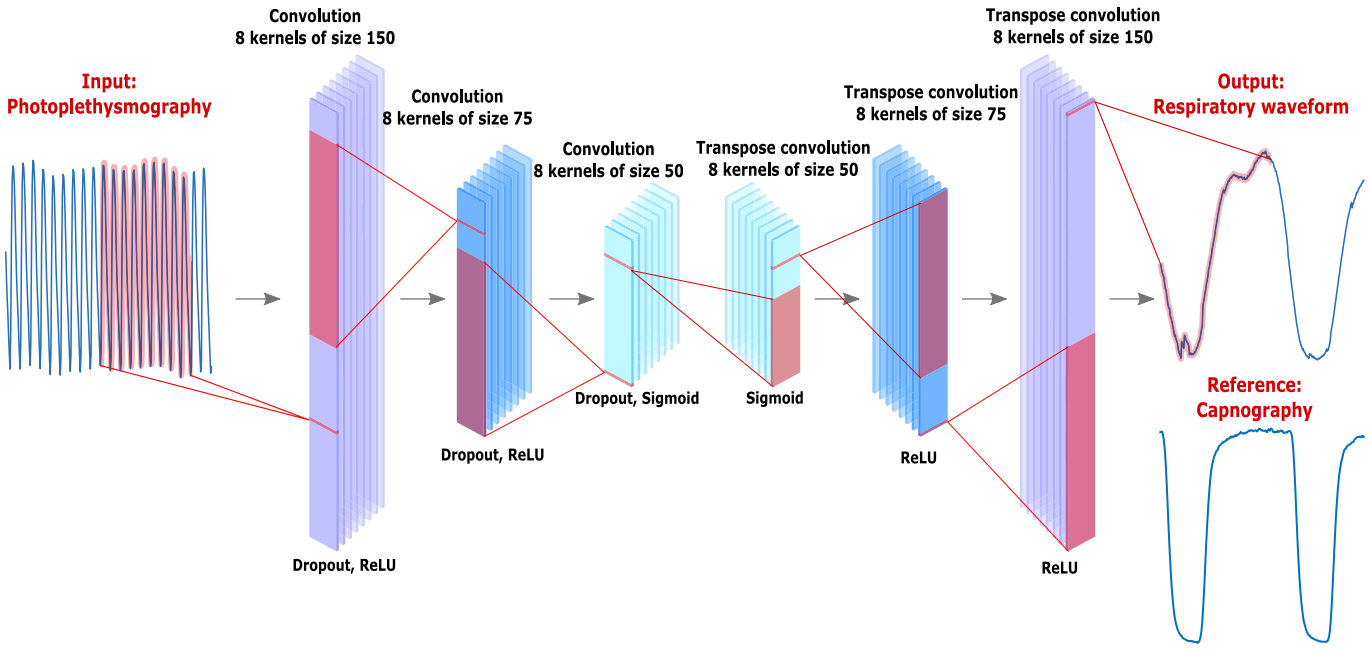

Fig. 1. The proposed deep learning architecture. From left to right: The photoplethysmography waveform is passed through 3 one-dimensional convolutional layers, each with 8 kernels. The resulting latent space is then passed through 3 transposed one-dimensional convolutional layers which up-sample the latent space into an output respiratory waveform. Under each layer the activation function (rectified linear unit or sigmoid) is labelled, along with the dropout. The output respiratory waveform is depicted next to the reference respiratory waveform, which would be used to calculate mean squared error during training. The waveforms shown in this figure are from a test subject example during leave-one-subject-out cross validation.

2.2. Deep learning model

A one-dimensional convolutional encoder–decoder structure was implemented in Pytorch [22]. An overview of the model structure is shown in Fig. 1. It consisted of 3 convolution "encoder" layers with 8 kernels each of sizes 150, 75 and 50 samples, which was then mirrored with 3 upsampling transpose convolution "decoder" layers. Both the PPG input and the output respiratory waveform estimate were 288 samples in length, corresponding to 9.6 s of data. The 9.6 s was simply a result of dividing the overall 480 s of data for each subject into 50 equal segments. Whilst this was somewhat arbitrary, it was important to have a window length that contained at least one full respiratory cycle. A rectified linear activation function was applied to the first two layers, with a sigmoid activation function being applied to the inner-most layers. A dropout of 0.5 was applied to the encoder layers. The input and output of the first layer were both padded by 20, and the output of the second layer was padded by 10. This was again mirrored in the transpose convolution layers within the decoder. The model was trained by minimising the mean square error between the output and the gold standard capnography. Adam optimisation was implemented with a learning rate of 1×10^{-3}.

The model parameters were optimised exclusively on the Capnobase data set, with training taking place on the first 39 subjects and testing on the last 3 subjects which had a good range of signal quality and breathing rates. When choosing the number of layers, there was a minor performance improvement when increasing from 2 encoder and decoder layers to 3 encoder and decoder layers, but 4 layers lead to no increase in performance and a tendency to over fit. When choosing the number of kernels per layer, it can be seen in Fig. 2(c) that there was no significant increase in median test mean absolute error when more than 8 kernels were used per layer, and there was an increase in the interquartile range indicating that the consistency of the model decreased beyond this point. When selecting the kernel sizes, performance was largely unaffected when changing the layer 1 kernel size between 100 and 150, but layer 1 kernel sizes beyond 150 lead to a significantly decreased performance. We opted to maintain larger kernel sizes in the input layer than is conventional so as to offer interpretability when compared with the sizes of the input and output signals. Importantly, performance benefited significantly from having kernel sizes that decreased in size with subsequent layers. Padding was kept minimal for improved computation time but still non-zero, as no padding at all lead to decreased performance. High dropout in the encoder layers improved performance on test data, as it improved the models ability to generalise to different inputs, but dropout had a significant detrimental effect on accuracy when applied to the decoder layers. An increased learning rate of 1×10^{-2} had difficulty converging, and a decreased learning rate of 1×10^{-4} took far longer to converge to a solution, and therefore a learning rate of 1×10^{-3} was found to be the sweet spot. Early stopping was also implemented by terminating training at 80 epochs to avoid overfitting. Whilst not particularly necessary, ReLU was implemented in the shallow layers considering that the output signal was normalised between 0 and 1, and lead to minor decreases in mean squared error. The addition of sigmoid activation functions in the inner layers had a marked improvement on stability as they prevented values from exploding, due to their bounded output property.

2.3. Model evaluation

The model was first tested on the Capnobase data set and leave-one-subject-out cross-validation was implemented. When testing on unseen subjects, instead of subjects' photoplethysmography data being divided into 50 unique segments, data was segmented with a sliding window of 9.6 s (288 samples) that was shifted 1 s each time. This resulted in 471 test segments for each subject. The model output segments were then fused and averaged, within a given evaluation window size, to give a smoother respiratory waveform estimate. The reasoning behind testing with overlapping windows was that given the low complexity of the model, a 9.6 s window of respiratory waveforms could be computed in a fraction of a millisecond. Averaging subsequent windows improved the robustness of the output waveforms with practically no extra cost. Moreover, in a real world implementation, the previous 9.6 s of respiratory information would be computed with a fast refresh rate in a similar fashion, rather than waiting for another 9.6 s to pass for

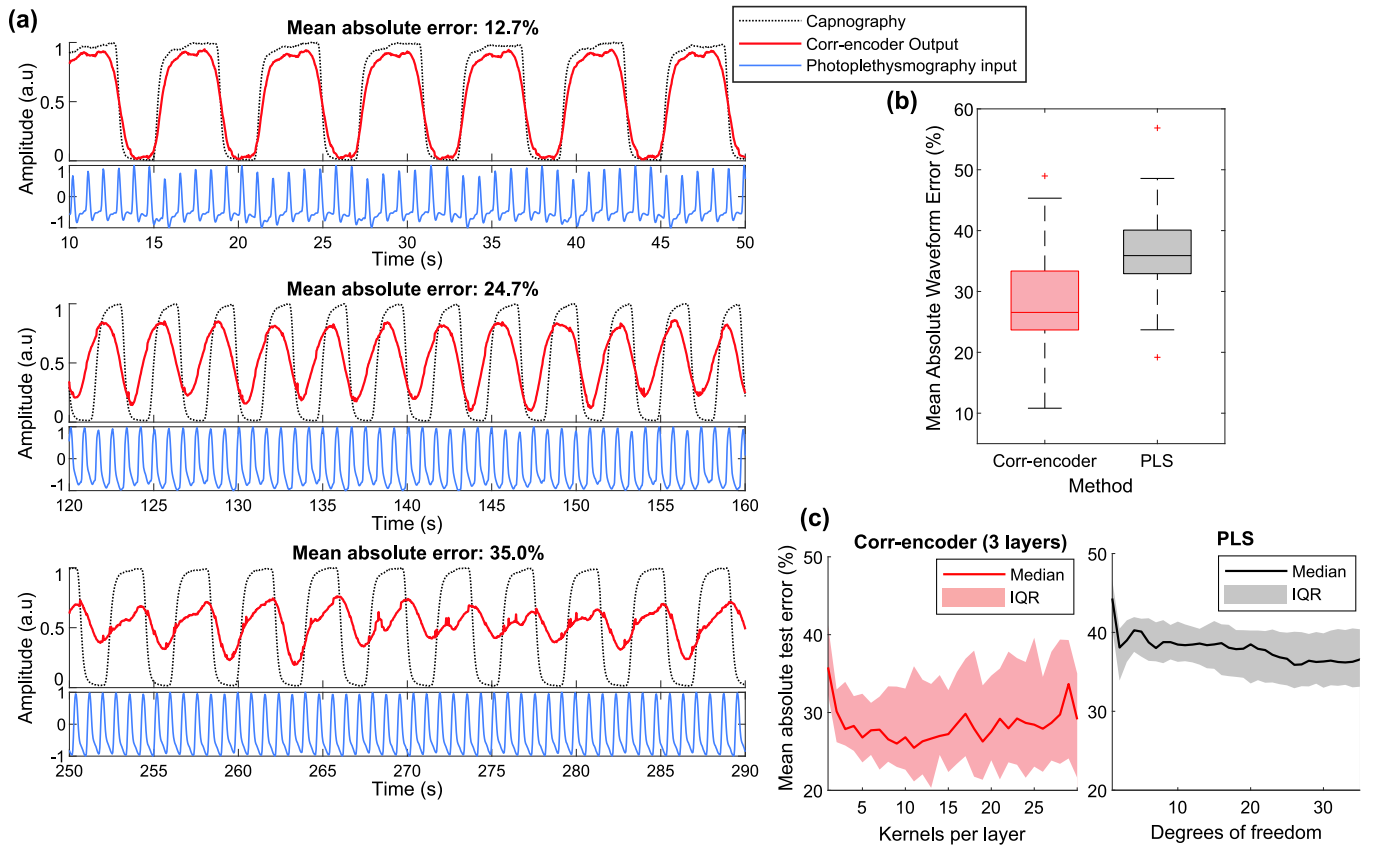

Fig. 2. Example output respiratory waveforms and mean absolute error resulting from the proposed model. (a) Output waveforms (red - solid), gold standard capnography waveforms (black - dotted) and input photoplethysmography waveforms (blue - solid) for three different scenarios (from top to bottom) of near perfect reconstruction, typical reconstruction and poor reconstruction. (b) Boxplots of test mean absolute waveform error across all 42 subjects of the Capnobase data set, for the proposed deep learning framework (red) and Partial Least Squares regression (grey). (c) Results of parameter sweeps for both the corr-encoder framework and PLS, showing the median and interquartile range of mean absolute waveform error across test subjects, against kernels per layer for the corr-encoder and degrees of freedom for PLS.

each new window. For these reasons, this strategy was implemented when testing the model. When training the model, the same strategy of creating overlapping inputs was unnecessary given the large volume of data that was available, and would have increased training time.

The output respiratory waveforms were first evaluated by comparing mean absolute error with the reference capnography waveform. The absolute mean error result was also compared to Partial Least Squares (PLS) regression [23], a standard way of finding projection onto latent spaces, which was trained and tested in exactly the same way using MATLAB. In this case, PLS was implemented with 26 degrees of freedom, which was found to give the optimal median mean absolute waveform error on test data. Partial Least Squares aims to predict an output from a given input by transforming an input through multiplication with input loading vectors, and transforming the output with output loading vectors. The algorithm is designed so that a paired transformed input has maximal covariance with its counterpart transformed output. In this sense, parallels can be drawn between the convolutional kernels of the corr-encoder model and the loading vectors of PLS. This comparison was important given that we are using the encoder–decoder structure in a similar way to what PLS was designed to accomplish. Partial Least Squares was evaluated in the same way to the deep learning model, through leave one subject out training and testing and averaging of overlapping test outputs. A parameter sweep for waveform mean absolute error against degrees of freedom is provided for PLS in Fig. 2(c), next to the corresponding analysis for the number of kernels per layer for the corr-encoder. It is interesting to note that the optimal total number of kernels for the corr-encoder of between 8 and 11 kernels per layer (resulting in 24 to 33 total encoder kernels) agrees with the optimal degrees of freedom of 26 for Partial Least Squares.

For the Capnobase data set, the output waveforms were evaluated for both respiratory rate estimation and for the duty cycle (the ratio between inspiration duration and expiration duration). To estimate the respiratory rate, the fast Fourier transform (FFT) was calculated and the frequency corresponding to the maximum power was extracted. This frequency in Hertz was then multiplied by 60 to give respiratory rate in breaths per minute. The respiratory rate estimation was compared to the gold standard respiratory rate, calculated manually in the data set from the capnography waveform. The FFT was chosen to evaluate our model over manual counting so that the respiratory rate calculation was automatic, making the results more realistic for real world implementation. Moreover, manual calculation was chosen above the FFT for comparison with the reference signal so that the reference respiratory rate could still be considered as a ground truth. To allow comparison with different models in the literature, both the median absolute respiratory rate error across all windows and the median mean absolute error across all subjects was calculated. For duty cycle estimation, given that the waveforms were trained to an amplitude between 0 and 1, the inspiratory duty cycle was estimated as the number of samples below 0.5 was divided by the total number of samples. This was then multiplied by 100 to give a percentage. The same operation was performed on the gold standard capnography for comparison. Both the mean and median absolute error in duty cycle are presented.

For the BIDMC data set, only respiratory rate estimation was used as an evaluation metric, given that the gold standard respiratory waveform was impedance pneumography and no longer capnography. Two training paradigms were implemented for evaluation. Firstly, the model was trained on the entirety of the Capnobase data set and then tested on

Table 1
Respiratory rate estimation results.

Authors	Dataset	Window size	Error method	Reported error (bpm)	Proportion of data used	Method
Davies (Proposed)	Capnobase	30.6s	Median AE	0.37 (0.19–1.42)	100%	Encoder–decoder with FFT
Khreis 2019 [24]	Capnobase	32s	Median AE	0.5 (0.2–1.1)	100%	Kalman Smoothing
Pimentel 2016 [20]	Capnobase	32s	Median Mean AE	1.5 (0.3–3.3)	92%	Spectral fusion
Davies (Proposed)	BIDMC	30.6s	Median AE	0.89 (0.36–3.05)	100%	Encoder–decoder with FFT
Aqajari 2021 [25]	BIDMC	30s	Mean AE	1.9 ± 0.3	100%	Generative adversarial networks
Pimentel 2016 [20]	BIDMC	32s	Median Mean AE	4.0 (1.8–5.5)	94%	Spectral fusion

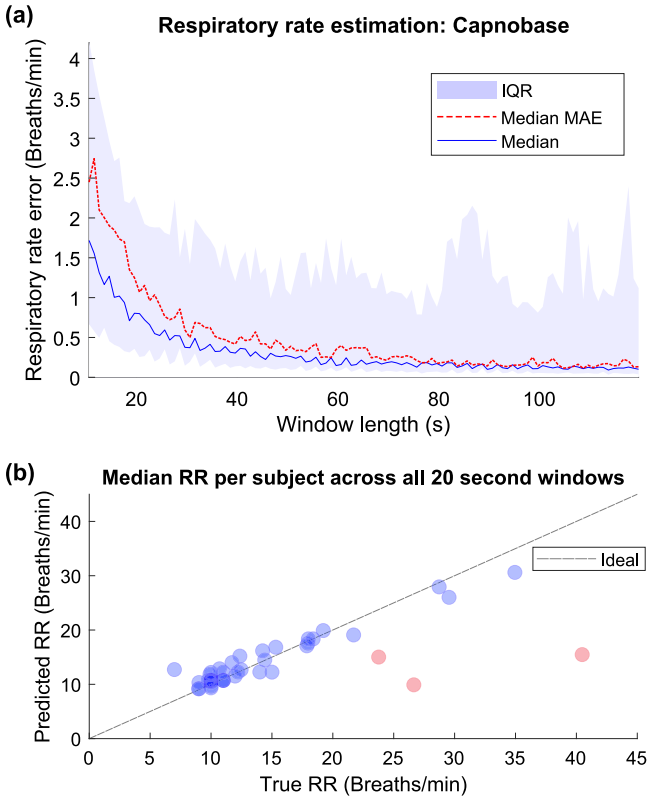

Fig. 3. Respiratory rate estimation results for the Capnobase data set. (a) Respiratory rate estimation error across all window lengths, with the median across all windows shown in blue, and the median mean absolute error across all subjects shown in red. (b) Scatter plot of the predicted median respiratory rate against true median respiratory rate for each subject, calculated using all 20 s estimation windows for a given subject. Three of the 42 subjects where the model performed poorly are highlighted in red.

the BIDMC dataset. Secondly, the model was trained on the Capnobase dataset, and then retrained on the BIDMC dataset in a leave one subject out fashion. In all cases in this work, the model was evaluated on unseen test subjects.

3. Results

3.1. Waveform estimation

Examplar test output waveforms are plotted against the gold standard capnography for three examples of good (MAE of 13%), typical (MAE of 25%) and poor (MAE of 35%) in Fig. 2(a). It can be seen that even when the difference between waveforms is large in the example with an MAE of 35%, the basic frequency components are captured. The median mean absolute waveform error of the model outputs across unseen test subjects was 27%, with an interquartile range of 24% to 33%. This compares to a median MAE of 36% and interquartile range of 33% to 40% for Partial Least Squares regression, as shown in Fig. 2(b).

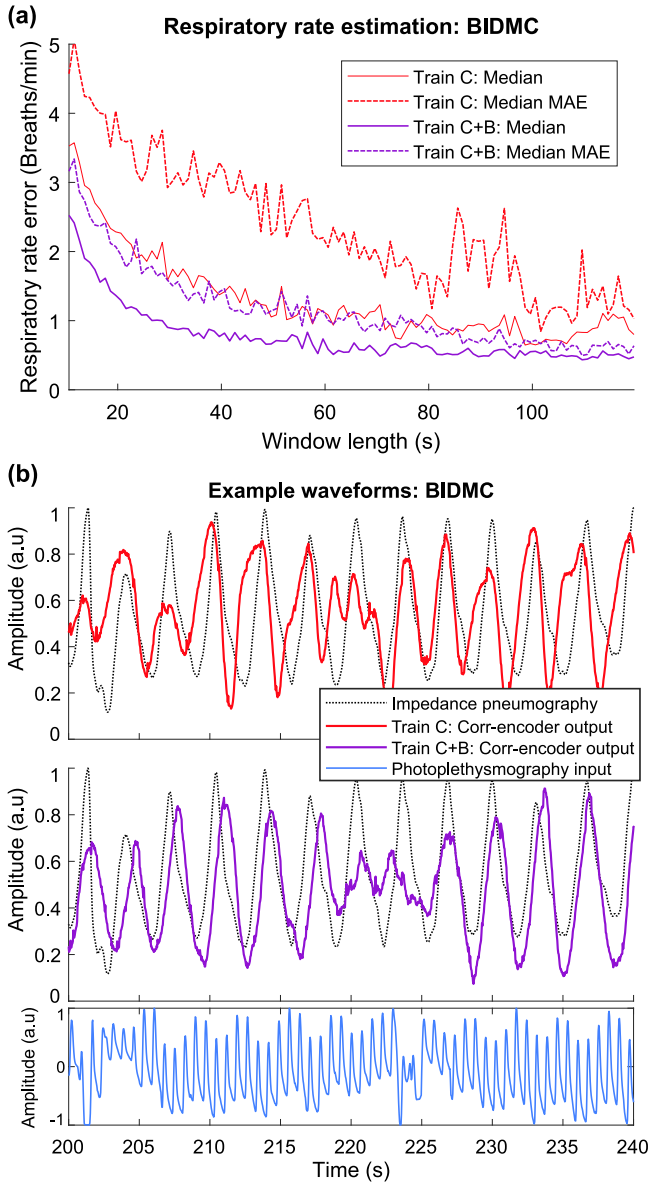

Fig. 4. Respiratory rate estimation results and example output waveforms for the BIDMC data set. (a) Median absolute error across all windows (solid line) and median mean absolute error across all subjects (dotted line) shown for training on Capnobase and testing on BIDMC (red) and training on Capnobase and then retraining on BIDMC through leave one subject out cross validation (purple). (b) Example output waveforms shown for training on Capnobase and testing on BIDMC (red) and training on Capnobase and then retraining on BIDMC through leave one subject out cross validation (purple), against the gold standard impedance pneumography (grey - dotted) and input photoplethysmography (blue).

3.2. Respiratory rate estimation

For comparability with other attempts at respiratory rate (RR) estimation in the literature, RR estimation error is presented in both median absolute error (mAE) across all windows, and median mean absolute error (mMAE) across all subjects. For windows of 30.6 s, the mAE was 0.37 breaths per minute, and the mMAE was 0.50 breaths per minute. For 60.6 s windows, these results improved slightly with a mAE of 0.16 breaths per minute and a mMAE of 0.40 breaths per minute. The MAE and mMAE are plotted for all time windows from 10 s to 120 s in Fig. 3(a). In Fig. 3(b), the median mean error is plotted for each subject across 20.6 s estimation windows, with a dotted line indicating the ideal prediction for comparison. This shorter window length was chosen to illustrate that even at window lengths far shorter than what has been presented thus far in the literature, respiratory rate estimation is broadly accurate across the full range of respiration frequencies and with no bias. It is further shown that estimation is poor in 3 out of the 42 subjects (highlighted in red), and this error persisted no matter what window size was implemented.

3.3. Respiratory timing

When comparing the estimated time spent breathing in, as a proportion of overall breathing cycle, between the output of the deep learning model and the gold standard capnography, there was median Pearson's correlation coefficient of 0.16 ($P. = 1.8 \times 10^{-3}$). This weak positive correlation indicates that for the most part, respiratory waveform estimations from the finger photoplethysmography failed to fully capture the timing ratio between inspiration and expiration.

3.4. Validation on BIDMC

When training the model on the Capnobase dataset, and testing on the unseen BIDMC dataset, respiratory rate estimation was broadly accurate with an MAE and mMAE of 1.74 and 3.02 breaths per minute at 30.6 s respectively, and an MAE and mMAE of 1.08 and 2.06 breaths per minute respectively, using 60.6 s windows. When training the model on the Capnobase dataset, and then retraining on the BIDMC dataset, these results were improved with an MAE and mMAE of 0.89, 1.50 and 0.58, 1.01 breaths per minute with 30.6 s windows and 60.6 s windows respectively. The respiratory rate estimate error is shown for all windows between 10 s and 120 s in Fig. 4(a). Moreover, it is shown in Fig. 4(b) that with a typical example waveform that with no training on BIDMC, the vast majority of the respiratory information is still captured. Retraining on BIDMC is shown to have the effect of changing the output waveform morphology to resemble that of the impedance pneumography which was in this case used as a reference, rather than capnography. For the purposes of comparison, the corr-encoder model was also trained exclusively on BIDMC, resulting in an MAE and mMAE of 0.90, 1.58 and 0.81, 1.36 breaths per minute with 30.6 s windows and 60.6 s windows respectively. These results were worse than if the model was pretrained on Capnobase and retrained on BIDMC, suggesting that the Capnobase dataset contains additional respiratory information that is not captured by BIDMC, and thus training on both sets of data enabled the model to better generalise to unseen subjects of BIDMC.

4. Discussion

For the reconstruction of capnography waveforms, the encoder–decoder structure presented here outperforms the standard Partial Least Squares regression. It is shown that the model succeeds in near perfect reconstruction of amplitude, phase and respiratory timing in only a small proportion of subjects. In the majority of cases the amplitude information is not captured completely, leading to mean absolute errors greater than 0.23 in three quarters of the test subjects. It can, however, be seen that in waveforms where amplitude information is not captured, resulting in a higher mean absolute waveform error, the dominant frequency information of the respiratory waveform is still effectively captured.

This is further reflected in the capability of the model for accurate estimation of respiratory rate, obtained simply by taking the fast Fourier transform of the output waveforms. Table 1 shows that when compared to other respiratory rate estimation algorithms, the waveforms produced by our model produce similar accuracy to the state of the art in both the Capnobase and BIDMC data sets. When forming these comparisons, care was taken to include recent examples which used all of the data in a given set, as well an algorithm that had been employed on both data sets. Comparisons with algorithms implemented on the BIDMC data set was particularly important, given the difficulty that it presents to respiratory rate estimation through the presence of dominant low frequency non-respiration related variations in the PPG [20]. Whilst our error was lower than Pimentel et al. [20] on the Capnobase dataset, it is important to note that Khreis et al. whom employed a method based on Kalman smoothing [24], achieved a comparable accuracy to our method on the Capnobase data set. The Kalman smoothing algorithm relies on a state transition matrix formed off the previous 32 s of data, which has the effect of smoothing out rapid fluctuations in breathing rate. Although this limitation does not impact accuracy on the Capnobase data set, where the breathing rate is generally steady over the course of each 8 minute recording, it would impact accuracy in applications with dramatic changes in breathing rate. This limitation does not apply to our proposed corr-encoder model, which can respond immediately to sharp changes in breathing rate. On the BIDMC data set, Aqajari et al. achieved comparable accuracy to our method with cycle generative adversarial networks (cycleGAN) [25]. The particular cycleGAN architecture implemented involved a generative network consisting of 16 convolutional layers and over 1600 kernels [26] and a discriminator network consisting of 4 layers and over 900 kernels [27], which when combined is a 20 layer model with over 2500 kernels. This is in contrast to our far simpler model, which had 6 layers with a combined 48 kernels in total. Given the comparatively low complexity of our architecture, it was possible to train our model in under 6 min on a CPU, versus the cycleGAN architecture which when applied to images can take hours to train on a GPU [26].

Respiratory waveform analysis that goes beyond respiratory rate, such as examining the ratio in timing between breathing in and breathing out (duty cycle), can be important for distinguishing respiratory diseases [1]. Successful estimation of the inspiratory duty cycle has previously been accomplished from in-ear PPG for the purposes of classifying chronic obstructive pulmonary disease [8]. Here we show that with the combination of the method presented in this work and the Capnobase data set, extraction of accurate duty cycle was largely unsuccessful, achieving only a weak positive correlation between the true duty cycle and our estimation. The low-pass filter transfer function between breathing and PPG [17] generally results in a loss of duty cycle detail [8]. Respiratory variations are generally weaker from finger PPG than they are from the ear-PPG [8,28], which could help to explain the lack of success of duty cycle estimation in this case.

It is important to note that other capnography waveform differences such as the saw tooth pattern, associated with obstructive breathing disorders, and rare CO_2 amplitude changes were not captured by this model. In the case of specific absolute values of CO_2 concentration in breath, it is fair to assume that this information would not be encoded within the PPG signal. In the case of the saw-tooth waveform however, it is feasible that relevant information is contained within photoplethysmography, and therefore in this case it likely was not captured by the model due to a lack of examples in the data set.

Currently, this trained model would be applicable in scenarios where finger photoplethysmography is recorded and is high-pass filtered to remove the DC offset. It is important to note that more data

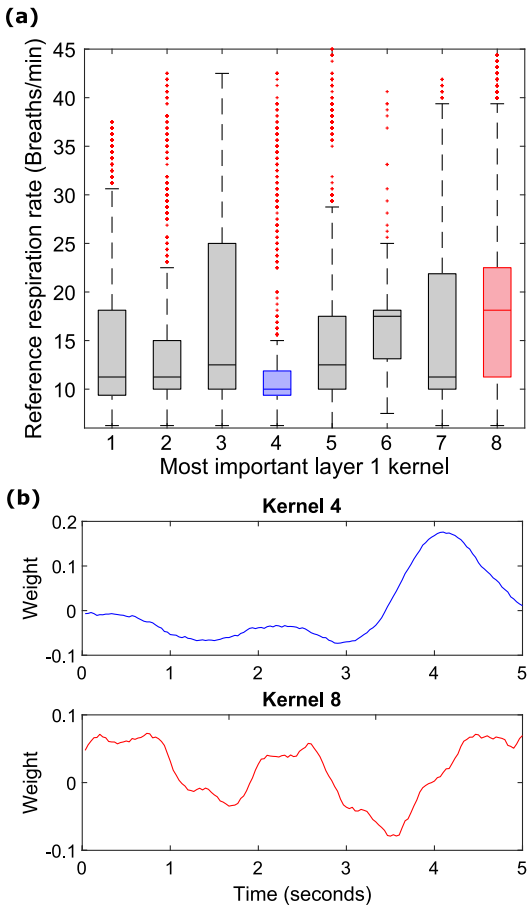

Fig. 5. Interpretability of kernel weights according to respiratory rate. (a) Boxplots of the reference respiratory rates for which a given layer 1 kernel produces the largest value in the latent space. The calculation of this is explained in depth in the "Model Interpretability" subsection of the Discussion. (b) Kernel weights for kernel 4 (blue) which has a strong affinity for low frequencies, and kernel 8 (red) which has a strong affinity for higher respiratory rates. Kernels weights in these plots are firstly smoothed with a moving average filter of 1 s in length.

would be required to train the model for other PPG locations, such as the ear, in which the PPG waveform is structurally different and more dominant respiratory variations are present [8]. Furthermore, the model has currently only been trained on data where patients were predominantly still, and thus to be effective in scenarios with lots of motion artefacts (such as during 6 minute walk tests), it is likely that more training data from these scenarios would be required.

The relatively low complexity of the deep learning model presented in this work resulted in ultra fast implementation speeds. The mean time for the model to compute a single 9.6 s respiratory waveform window is 0.43 ms on an Intel i7-1165G7 processor (2.8 GHz), and therefore over 6 h worth of respiratory waveforms can be generated in under 1 s. This compares to 5 min of data per second when using Kalman smoothing of different frequency estimates (2.4 GHz) [20] and 4 min of data per second when using an empirical mode decomposition approach combined with independent component analysis (3.2 GHz) [29]. Practically, the speed of our proposed method would allow for easy implementation on small form factor microcontrollers in wearable devices.

4.1. Model interpretability

Especially for health care applications, it is important to attempt to dispel the black box of deep learning methods. Given the relatively low complexity of this model, and convolutional kernels that are comparable in length to the input, we can start to infer what the model is selecting for in the input data. To calculate this, trained weights for the first 3 layers of the model were extracted. For all input windows of photoplethysmography from the Capnobase data set, convolution was performed with all different possible combinations of the ordered convolutional kernels. For 8 different kernels in each layer, this resulted in 512 different possible projections of a given input to the latent space. Whilst the majority of combinations resulted in zeros and negligible values, some projection paths for a given input resulted in large latent values. The maximum across all 512 combinations was then calculated, and linked to the specific kernel in layer 1 which triggered this maximum value. This allows us to infer the most important layer 1 kernel weights for a given input. The most important layer 1 kernels were then plotted against the corresponding reference respiratory rate for every input, resulting in the boxplots shown in Fig. 5(a). Given that the chosen kernel length was of similar size to the input, it can be seen that in some part these kernels are discriminating the different inputs based on frequency. This deduction is further reinforced when we visualise the smoothed weights of kernel 4 (triggering predominantly at low frequency respiration with a median reference respiratory rate of 10 breaths per minute) and kernel 8 (triggering predominantly at higher frequencies with a median reference RR of 18 breaths per minute) in Fig. 5(b). The weights of kernel 4 correspond to a low frequency pattern of respiration, whereas the weights of kernel 8 contain higher frequency information. This result is important as it shows that the input model is in part extracting frequency based information from the input, in a similar way to performing a discrete cosine transform. This knowledge helps us to move from a black box interpretation of the model towards a grey box.

5. Conclusion

We have introduced a novel low complexity deep learning framework for the extraction of respiratory information from photoplethysmography. The proposed model has been shown to produce accurate respiratory waveforms and state of the art respiratory rate estimates. We have demonstrated this effectiveness on two different data sets, Capnobase and BIDMC, highlighting its strong ability to generalise. Importantly, given the relative simplicity of our model, it has been shown to be capable of transforming several hours of photoplethysmography data into corresponding respiratory waveforms in a single second. The framework presented in this work helps to bridge the gap between photoplethysmography and gold standard respiratory references such as capnography and impedance pneumography. We believe that our model can therefore improve the usefulness of consumer PPG-based wearables in medical applications.

CRediT authorship contribution statement

Harry J. Davies: Conceptualization, Methodology, Software, Validation, Writing, Visualization. **Danilo P. Mandic:** Conceptualization, Writing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

We declare no conflicts of interest for the work in this article.

Data availability

The data sets used in this paper are publically available and code will be shared upon request.

Acknowledgements

This work was supported by the Racing Foundation grant 285/2018, MURI/EPSRC grant EP/P008461, and the Dementia Research Institute at Imperial College London.

References

[1] M.J. Tobin, T.S. Chadha, G. Jenouri, S.J. Birch, H.B. Gazeroglu, M.A. Sackner, Breathing patterns: 2. Diseased subjects, Chest 84 (3) (1983) 286–294.

[2] H. Wilkens, B. Weingard, A. Lo Mauro, E. Schena, A. Pedotti, G.W. Sybrecht, A. Aliverti, Breathing pattern and chest wall volumes during exercise in patients with cystic fibrosis, pulmonary fibrosis and COPD before and after lung transplantation, Thorax 65 (9) (2010) 808–814.

[3] D.J. Meredith, D. Clifton, P. Charlton, J. Brooks, C.W. Pugh, L. Tarassenko, Photoplethysmographic derivation of respiratory rate: A review of relevant physiology, J. Med. Eng. Technol. 36 (1) (2012) 1–7.

[4] K.H. Shelley, A.A. Awad, R.G. Stout, D.G. Silverman, The use of joint time frequency analysis to quantify the effect of ventilation on the pulse oximeter waveform, J. Clin. Monit. Comput. 20 (2) (2006) 81–87.

[5] L. Nilsson, A. Johansson, S. Kalman, Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique, J. Clin. Monit. Comput. 16 (4) (2000) 309–315.

[6] N. Rehman, D.P. Mandic, Multivariate empirical mode decomposition, Proc. R. Soc. A 466 (2117) (2010) 1291–1302.

[7] M.A. Motin, C.K. Karmakar, M. Palaniswami, Selection of empirical mode decomposition techniques for extracting breathing rate from PPG, IEEE Signal Process. Lett. 26 (4) (2019) 592–596.

[8] H.J. Davies, P. Bachtiger, I. Williams, P.L. Molyneaux, N.S. Peters, D.P. Mandic, Wearable in-ear PPG: Detailed respiratory variations enable classification of COPD, IEEE Trans. Biomed. Eng. 69 (7) (2022) 2390–2400.

[9] P.H. Charlton, D.A. Birrenkott, T. Bonnici, M.A.F. Pimentel, A.E.W. Johnson, J. Alastruey, L. Tarassenko, P.J. Watkinson, R. Beale, D.A. Clifton, Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review, IEEE Rev. Biomed. Eng. 11 (2018) 2–20.

[10] M.S. Siobal, Monitoring exhaled carbon dioxide, Respir. Care 61 (10) (2016) 1397–1416.

[11] H. Aminiahidashti, S. Shafiee, A. Zamani Kiasari, M. Sazgar, Applications of end-tidal carbon dioxide (ETCO2) monitoring in emergency department: A narrative review, Emerg. (Tehran) 6 (1) (2018) e5.

[12] A. Gogna, A. Majumdar, R. Ward, Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals, IEEE Trans. Biomed. Eng. 64 (9) (2017) 2196–2205.

[13] D.-H. Kang, D.-H. Kim, 1D convolutional autoencoder-based PPG and GSR signals for real-time emotion classification, IEEE Access 10 (2022) 91332–91345.

[14] L. Gondara, Medical image denoising using convolutional denoising autoencoders, in: Proc of the 16th IEEE International Conference on Data Mining Workshops (ICDMW), 2016, pp. 241–246.

[15] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, S.-Y. Chien, Noise reduction in ECG signals using fully convolutional denoising autoencoders, IEEE Access 7 (2019) 60806–60813.

[16] W. Karlen, S. Raman, J.M. Ansermino, G.A. Dumont, Multiparameter respiratory rate estimation from the photoplethysmogram, IEEE Trans. Biomed. Eng. 60 (7) (2013) 1946–1953.

[17] A. Johansson, P.A. Öberg, Estimation of respiratory volumes from the photoplethysmographic signal. Part 2: A model study, Med. Biol. Eng. Comput. 37 (1) (1999) 48–53.

[18] R. Katiyar, V. Gupta, R.B. Pachori, FBSE-EWT-based approach for the determination of respiratory rate from PPG signals, IEEE Sens. Lett. 3 (7) (2019) 1–4.

[19] M.A. Motin, C.K. Karmakar, M. Palaniswami, Ensemble empirical mode decomposition with principal component analysis: A novel approach for extracting respiratory rate and heart rate from photoplethysmographic signal, IEEE J. Biomed. Health Inf. 22 (3) (2018) 766–774.

[20] M.A.F. Pimentel, A.E.W. Johnson, P.H. Charlton, D. Birrenkott, P.J. Watkinson, L. Tarassenko, D.A. Clifton, Toward a robust estimation of respiratory rate from pulse oximeters, IEEE Trans. Biomed. Eng. 64 (8) (2016) 1914–1923.

[21] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet, Circulation 101 (23) (2000) e215–e220.

[22] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035.

[23] S. de Jong, SIMPLS: An alternative approach to partial least squares regression, Chemometr. Intell. Lab. Syst. 18 (3) (1993) 251–263.

[24] S. Khreis, D. Ge, H.A. Rahman, G. Carrault, Breathing rate estimation using Kalman smoother with electrocardiogram and photoplethysmogram, IEEE Trans. Biomed. Eng. 67 (3) (2019) 893–904.

[25] S.A.H. Aqajari, R. Cao, A.H.A. Zargari, A.M. Rahmani, An end-to-end and accurate PPG-based respiratory rate estimation approach using cycle generative adversarial networks, in: Proceedings of the 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2021, pp. 744–747.

[26] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: Computer Vision – ECCV 2016, Springer International Publishing, 2016, pp. 694–711.

[27] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017.

[28] K. Budidha, P.A. Kyriacou, In vivo investigation of ear canal pulse oximetry during hypothermia, J. Clin. Monit. Comput. 32 (1) (2018) 97–107.

[29] R. Lei, B.W.-K. Ling, P. Feng, J. Chen, Estimation of heart rate and respiratory rate from PPG signal using complementary ensemble empirical mode decomposition with both independent component analysis and non-negative matrix factorization, Sensors 20 (11) (2020) 3238.