Respiratory Rate Estimation using PPG: A Deep Learning Approach

Dayi Bian, Pooja Mehta, and Nandakumar Selvaraj*

Abstract— **Respiratory rate (RR) is an important vital sign marker of health, and it is often neglected due to a lack of unobtrusive sensors for objective and convenient measurement. The respiratory modulations present in simple photoplethysmogram (PPG) have been useful to derive RR using signal processing, waveform fiducial markers, and hand-crafted rules. An end-to-end deep learning approach based on residual network (ResNet) architecture is proposed to estimate RR using PPG. This approach takes time-series PPG data as input, learns the rules through the training process that involved an additional synthetic PPG dataset generated to overcome the insufficient data problem of deep learning, and provides RR estimation as outputs. The inclusion of a synthetic dataset for training improved the performance of the deep learning model by 34%. The final mean absolute error performance of the deep learning approach for RR estimation was 2.5±0.6 brpm using 5-fold cross-validation in two widely used public PPG datasets (n=95) with reliable RR references. The deep learning model achieved comparable performance to that of a classical method, which was also implemented for comparison. With large real-world data and reference ground truth, deep learning can be valuable for RR or other vital sign monitoring using PPG and other physiological signals.**

Index Terms—**Photoplethysmography, Respiratory Rate, Deep Learning, Convolutional Neural Network**

I. INTRODUCTION

Respiratory rate (RR) is a prominent diagnostic marker of respiratory dysfunction and dysfunction and this crucial vital sign is also commonly known as breathing rate or breathing frequency. Abnormally elevated RR is a good predictor for cardiac arrest and highly correlated with in-hospital mortality. Consequently, monitoring of RR is fundamental to assess patient's health status in hospital and home or community settings. However, the current clinical practice usually measures RR by counting chest wall expansion, which is highly subjective, prone to inaccuracy, and inconvenient. Unobtrusive continuous measurement of RR has been evolving using any or a combination of impedance pneumography (IP), piezoelectric, acceleration, electrocardiogram (ECG), and photoplethysmogram (PPG) sensors applied for research and patient monitoring applications. RR monitoring is not available in widespread wearable fitness sensors and smart watches, despite other key physiological measurements, including heart rate (HR), heart rate variability, and temperature, are more accessible and useful for clinical purposes. Therefore, there is a tremendous value in obtaining reliable RR estimation using biosensor medical devices and convenient wearable sensors.

A wide range of existing wearable sensors and smart watches record PPG signal at one or more wavelengths of green, red and infrared measuring relative changes in blood perfusion volume of a local tissue over time. The PPG is modulated predominantly by respiratory system among other numerous physiological mechanisms and manifested as three types of signal modulation: baseline wander (BW) of the PPG that is influenced by the changes in intrathoracic pressure and vasoconstriction of arteries during inhalation; amplitude modulation (AM) of the PPG that reflects the changes in stroke volume and intrathoracic pressure during respiratory cycles; and frequency modulation (FM) is nothing but the respiratory sinus arrhythmia that exhibits HR to increase during inspiration and to decrease during expiration. Consequently, RR can be estimated from these respiratory surrogate signals derived from PPG.

A comprehensive research has been established for PPG-based RR estimation [1]. RR algorithms commonly involve digital filtering, time and or frequency domain analysis, signal decompositions, the fiducial points derived respiratory surrogate waveforms and features, signal quality estimates and sensor fusion [2]. The above classical approaches rely on hand-crafted rules and empirical parameters optimized for specific algorithmic methods, designed for general or specific target patient population.

In contrast, deep learning has been recently explored for numerous biometric analytics involving PPG time-series signals. Convolutional neural networks (CNN) have been trained for detecting atrial fibrillation using wavelet transforms of PPG combined with hand-crafted features as inputs [3] and for estimating HR using the spectrogram of PPG and accelerometer signals as inputs [4]. Biswas et al. [5] used PPG data to train a deep neural network combining a CNN and an recurrent neural network (RNN) for biometric based personal identification and HR estimation.

One of the limitations of deep learning approach is that it requires a relatively large amount of annotated data. Given the limited datasets with RR annotations are publicly available, generating synthetic data is an effective approach to mitigate the requirement of a large amount of annotated data for deep learning approach [6]. The periodic PPG time-series data have been synthesized for the comparison of performances of various RR estimation algorithms [2]. Selvaraj et al. [7] also generated a synthetic ECG dataset, which shares some similar principles to that of synthetic PPG generation, for the performance validation of RR estimation.

An end-to-end deep learning approach based on convolutional neural network architecture is currently proposed for RR estimation using PPG. Accordingly, the deep learning architecture takes raw PPG data as input, learns the rules through training, and estimates RR values as outputs. The

Authors are with Biofourmis Inc., Boston, MA 02110 USA. (*correspondence e-mail: kumar@biofourmis.com)

deep learning model is trained by adding real-world data, synthetic data and augmented data to the training dataset systematically as well as for different combinations. The systemic improvement in performances for RR estimation are assessed and also compared to that of a classical method implemented additionally for RR estimation. The remaining paper is organized as follows: Section II introduces the datasets used and describes both the classical method and the deep learning-based methods. Section III summarizes the results of our proposed RR estimation methods. The final Section concludes the paper with limitations and future work.

II. MATERIALS AND METHODS

A. Datasets

A synthetic PPG dataset is currently generated to mitigate the requirement of a large amount of annotated data for deep learning. In addition to the synthetic simulated data, two publicly available datasets, CapnoBase and BIDMC, have been employed for training and cross validation of deep learning model. Furthermore details are given below.

1) Synthetic dataset: Synthetic PPG data were generated as follows. Single PPG cycle templates are controlled to generate trains of PPG beats with appropriate cycle durations determined by the desired input HR. Each synthetic PPG record is produced for a predetermined time duration (e.g., 60s) and sample rate (e.g., 300 Hz). All the respiratory modulations (AM, FM, and BW) driven by the desired input RR are independently added to the PPG data to generate modulated PPG data (Fig. 1). Furthermore, the intensities of these respiratory modulations are varied and Gausssian white noise is added to the synthetic data in order to mimic realistic variability and inherent noise scenario. The HR and RR inputs for synthetic data generation were randomly chosen from the predetermined ranges of 30 to 200 beats-per-min (bpm) and 4 to 60 breaths-per-min (brpm), respectively. Taking the Nyquist criterion and physiological variability into consideration [7], a predetermined number of physiologically possible pairs or combinations of HR and RR from their respective range are chosen for the generation of PPG recordings for the synthetic dataset. In the end, a total of 100,000 recordings are generated for the current example of deep learning model training.

2) CapnoBase dataset: This dataset [8] contains PPG recordings and capnography data, both sampled at 300 Hz. The cases in the dataset were randomly selected from a larger collection of physiological signals collected during elective surgery and routine anesthesia. The dataset consists of 42 recordings of 8-minute duration from 29 pediatric and 13 adult patients containing quality recordings under spontaneous and controlled breathing. The gold standard capnogram waveforms of the database have been manually labelled for each breath cycle by a research assistant, and the annotations were used to calculate the reference RR values based on the time between consecutive breaths.

3) BIDMC dataset: This dataset is extracted from the MIMIC-II resource [9] and comprised of PPG recordings and simultaneous IP respiratory signals from 53 adult intensive care patients recorded for about 8-minute duration both at a sampling rate of 125 Hz. The IP waveforms of each record were used as the reference respiratory ground truth, where each breath cycle in the IP signals is manually annotated by two research assistants independently, and both sets of annotations are used to calculate the reference RR values.

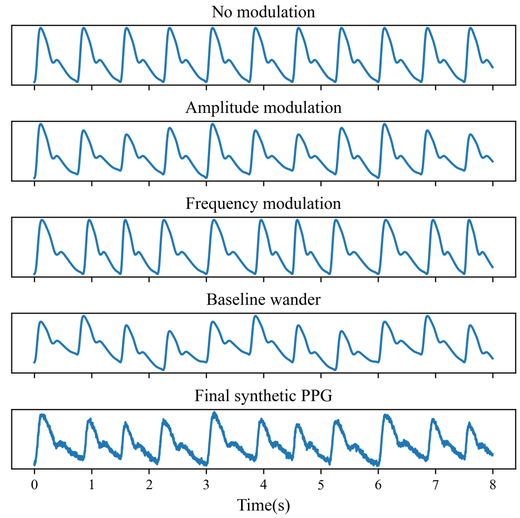

Fig. 1. Synthetic data from top to bottom: 1) with no respiratory modulation, with 2) amplitude modulation, 3) frequency modulation, and 4) baseline wander, and 5) the final synthetic PPG data adding all three respiratory modulations and additional white gaussian noise.

B. RR estimation methods

The optimal window size for RR estimation can range from 30s to 90s [10], where lower errors have been reported at longer window sizes [8], and a shorter window size, on the other hand, lowers the computational cost with high stability for RR estimation algorithms. A window size of 60s with 1s forward shift has been chosen currently for segmentation of the datasets that results in 100,000 synthetic PPG segments and 39,995 real PPG segments for RR estimation.

1) Classical method: A classical RR estimation method has been implemented including a fusion of independent RR estimates from multiple respiratory signals, as shown in Fig. 2. More details of such traditional algorithms can be found in [1]. Briefly, the raw PPG signal was down sampled and filtered to remove very low frequency components of the PPG. Then a peak detection method from the Python PeakUtils package was used to detect fiducial peaks for each beat with empirical parameters tuned for the PPG signal. The respiratory signals were extracted using feature-based techniques: 1) BW was extracted as the mean amplitude between the peaks and preceding troughs; 2) AM was extracted as the difference between the amplitudes of peaks and proceeding troughs; and 3) FM was extracted as the inter-peak intervals. After extracting the above surrogate respiratory signals, a respiratory peak detection algorithm was used to estimate the independent RRs from three surrogates signals. Standard deviation (SD) of RR estimations from respiratory surrogate signal segments was calculated and the independent RR estimates were set to be invalid if SD is larger than a

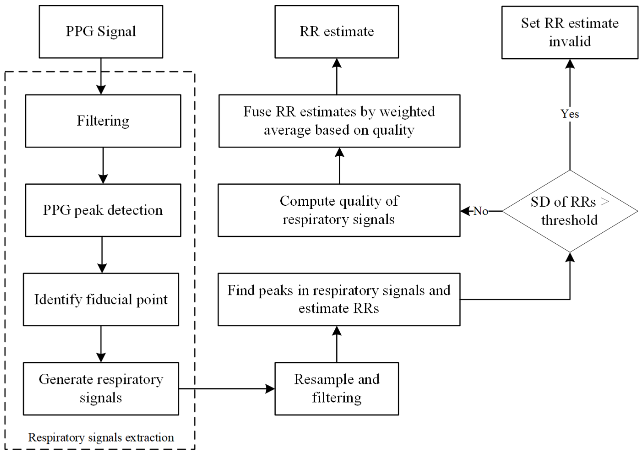

Fig. 2. Flowchart of the SmartQualityFusion RR estimation method.

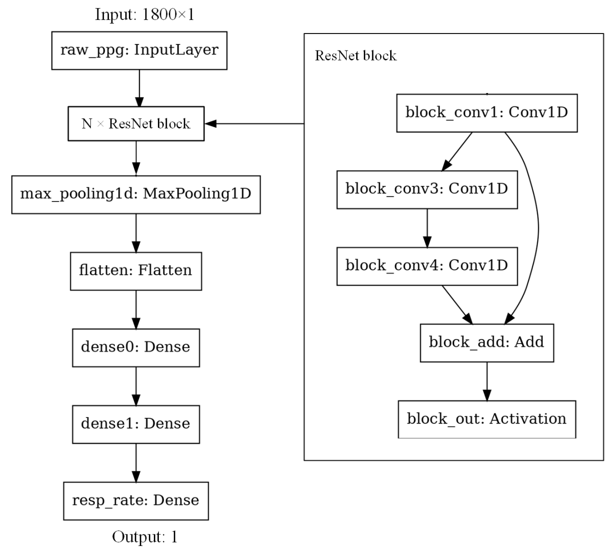

Fig. 3. Deep neural network architecture.

predetermined threshold. The intuition is that the RR outputs from different respiratory signals should not vary too much. After we examined the variations of the RR estimates, a quality metric ranging from 0 to 1 [11] was calculated for each respiratory signal, and the final RR was computed as the weighted mean of the RR from all three respiratory signals. This method is referred to as the SmartQualityFusion, as a combination of Smart fusion [8] and Quality fusion [11].

The public datasets were processed by the implemented classical RR estimation algorithm and produce the RR estimates for all 60s PPG segments in each record. Meanwhile, the instantaneous ground truth RR values were averaged to produce the reference RR values for each 60s segment. Mean absolute error (MAE) was calculated as a performance metric in each record by averaging the absolute error between the RR estimations from the classical algorithm and the ground truth reference RR values, and as given below:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|RR_{est}^{i} - RR_{ref}^{i}\right| \tag{1}$$

where N is total number of PPG segments, RR_{est}^{i} and RR_{ref}^{i} denote the estimated and ground truth RR for the ith segment respectively.

2) Deep learning method: In this study, a convolutional neural network (CNN) architecture is input with the continuous periodic PPG time series signal for the determination of RR as an output measure. To be more specific, a popular variant of CNN, namely residual network (ResNet), is currently employed for the measurement of RR using PPG signals. ResNet [12] inserts shortcut connections into the convolution layers which turn the CNN network into its counterpart residual version. Empirical evidence has shown that ResNets are easier to optimize and can gain performance boosts with increased network depth while still maintaining lower computational complexity, which can be substantially beneficial for real-time respiratory rate monitoring.

In the current study, the one dimensional raw PPG signal was resampled to 30Hz prior to feeding as an input to the ResNet model that preserves the PPG signal integrity well and substantially reduce the computational requirements and model complexity. The ResNet deep nueral network architecture is presented in Figure 3.

The deep learning network takes raw PPG data as an input and the initial convolution layer serves to filter the raw signal. The subsequent convolution layers increase the learning capabilities of the model and construct features throughout the layers. Rectified linear unit (ReLU) is used as the activation function throughout the entire network except for the output layer to introduce nonlinearity into the model. The last dense layer outputs the RR estimation as a continuous value. The loss function is defined as the mean absolute error (MAE) (Equation (1)) between the estimated RR and the ground-truth RR of the input PPG segment.

Selection of optimal hyperparameters is crucial for deep learning model development, since it ensures the proper model complexity and optimal learning structure for the given problem. The Bayesian optimization algorithm [13], which has shown success in hyperparameter tuning among various machine learning models, was used to select the optimal hyperparameters for the network. These hyperparameters include the number of ResNet blocks (N_{res}), filter size ($filter_c$), kernel size ($kernel_c$) for convolution layer, stride size for convolution layer ($stride_c$) and max pooling layer ($stride_p$), size of first dense layer (n_{den}^{1}), etc. Table I summarizes the hyperparameters and their values tested in Bayesian optimization. The Adam optimization method [14] is currently used to optimize this regression problem.

The proposed deep neural network is trained systematically by adding different datasets such as real-world data, synthetic data and augmented dataset to the training dataset one-by-one or in different combinations to assess the systemic improvement in performance for the estimation of RR. In order to find the best strategy to train the deep learning model using the above datasets, the following unique approaches are adapted to train the deep learning models:

- Use real data only to train models (DL:R).
- Use synthesized data to train the models first, and later

TABLE I

HYPERPARAMETERS FOR BAYESIAN OPTIMIZATION

Hyperparameters for tuning	**Range of values**
Number of ResNet blocks	2 ...6
Filter size of convolution layer	5 ...9
Kernel size of first convolution layer in ResNet block	3 ...5
Stride size of first convolution in ResNet block	2 ...5
Stride size of max pooling layer	2 ...5
Number of units in first dense layer	15 ...30
Whether use BatchNormalization	True, False
Learning rate for optimization function	1e-3, 1e-4, 1e-5

adapt the models with real data, similar to the idea of transfer learning.

- Train models with synthetic dataset with the intention to use them on real dataset. (DL:S)
- Fine tune the baseline model from DL:S with the real dataset. (DL:S→R)

- Use hybrid datasets including synthetic, real and augmented datasets for better model training.
  - Train models with hybrid datasets comprised of real dataset and synthetic dataset together. (DL:R+S)
  - With real and synthetic datasets, add inverted real PPG data to further augment the training data. (DL:R+S+A)

Due to the small sample size of real datasets, group fold cross-validation was carried out while evaluating our deep learning models involving real datasets. The real datasets were randomly split into five folds while any subject's data only appeared in 1 fold. Four folds are used for training and validation (3 folds for training, 1 fold for validation), while the remaining fold is used for testing. Thus, the training is performed in total 5 times, so that each fold serves as test data only once. Synthetic data and augmented data were also added to the training dataset as needed for different training approaches. The deep learning models were implemented in TensorFlow version 2.0. Training and evaluation were done on Nvidia Tesla V100 GPU with 16 GB of RAM hosted by AWS EC2 GPU instance. For each experiment, 100 epochs with a batch size of 128 were performed. The early stopping technique was used during training stage to reduce overfitting. The model with the lowest validation loss was chosen and used for testing.

III. RESULTS

The hyperparameters for ResNet deep neural network optimized by Bayesian optimization are presented in Table II. The final neural network comprises of five ResNet blocks following by one MaxPooling layer, one Flatten layer, and three Dense layers with decreased number of units. Each ResNet block contains three Convolutional layers with optimized kernel size, filter size, and stride size as listed in the table, as well as one Merging layer and one Activation layer.

The performances of RR estimation in MAE from cross-validation in real datasets are presented in Figure 4 and Table III. The results show that the models trained with only real data (DL:R) and synthetic data (DL:S) show a MAE of 3.8±0.5 brpm and 4.2±0.5 brpm, respectively.

TABLE II

OPTIMIZED HYPERPARAMETERS FOR DEEP NEURAL NETWORK

Layer		**Details**
ResNet blocks (i=1...5)	1^{st} convolution	$kernel_c$=3, $stride_c$=2, $filter_c^i$=6 if i=1 else $2 \times filter_c^{i-1}$
	2^{nd} convolution	Same as convolution 1, except $stride_c$=1
	3^{rd} convolution	Same as convolution 2
	Merge	Add function used
	Activation	ReLU
Max pooling		$stride_p$=2
Flatten		
1^{st} dense layer		n_{den}^1=20
2^{nd} dense layer		n_{den}^2=10
3^{rd} dense layer		n_{den}^3=1

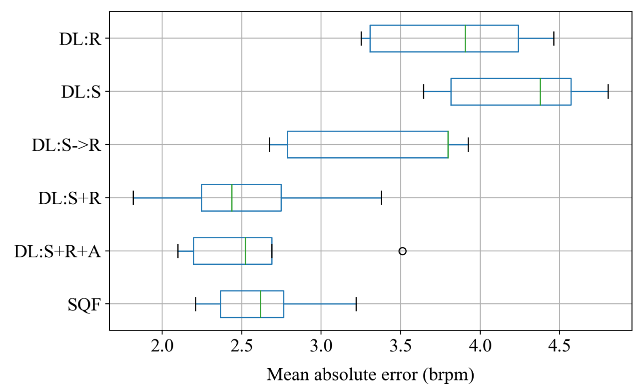

Fig. 4. Boxplot comparison of mean absolute error for deep learning methods and SmartQualityFusion method.

Fine tuning of the synthetic data based model with real data (DL:S→R) decreases the MAE from 4.2±0.5 brpm to 3.4±0.6 brpm. Thus, the transfer learning with synthetic data training followed by fine tuning with real world data showed an 11% improvement in MAE from 3.8 brpm to 3.4 brpm.

In contrast, when the synthetic data and real data are combined together (DL:S+R) as training data, the MAE of the model dramatically decreases from 3.8±0.5 brpm to 2.5±0.6 brpm leading to 34% improvement in MAE as shown in Table III. Appending the augmented data to the training dataset did not have any impact for further improvement in the performance of the model. The best performance result from the deep learning approach (DL:S+R, MAE = 2.5±0.6 brpm) is found to be comparable to that of the classical time-domain method (SQF) (MAE = 2.6±0.4 brpm) in these real datasets. The relationship between the reference and estimated RR from the best deep learning model is strong with a high correlation (r=0.96).

IV. DISCUSSION

Abnormal respiration is the first clinical sign of health deterioration. The RR vital sign is often neglected and seldom measured in and out of both clinical environment. Wearable devices integrated with simple PPG sensor have been increasingly adapted into clinical medicine due to their pervasive convenience and simplicity, which also provides

TABLE III
SUMMARY OF THE MAE PERFORMANCE (AVERAGE±STANDARD DEVIATION FOR ALL METHODS

Method		MAE (brpm)
Deep learning approach	Train with real data only (DL:R)	3.8±0.5
	Train with synthetic data only (DL:S)	4.2±0.5
	Fine tune the model in DL:S with real data (DL:S→R)	3.4±0.6
	Train with synthetic data and real data (DL:S+R)	2.5±0.6
	Train with synthetic data, real data, and inverted data (DL:S+R+A)	2.6±0.6
SmartQualityFusion method (SQF)		2.6±0.4

the opportunity to bridge the gap for objective and unobtrusive vital sign measurements. Current PPG based RR estimation methods rely heavily on hand-craft rules and parameters tuned for specific settings. The study presented a novel end-to-end learning approach based on deep learning (ResNets) to estimate RR from PPG, and demonstrated clinically reasonable performance.

Proper training of deep learning model requires relatively large amount of PPG data with RR annotations. Standard respiratory signals such as capnograms are obtrusive and less accessible, and generating RR annotations are laborious and expensive. In order to mitigate such problem, a synthetic PPG dataset was effectively generated, in the current study, to facilitate the deep learning model training process and the preliminary performance results are promising.

As compared to the baseline model trained without any synthetic data, the cross-validation results show that augmenting the real dataset with synthetic dataset enhanced the MAE from 3.8 brpm to 2.5 brpm. Thus, the proposed training strategies of deep learning model with synthetic data generation demonstrated substantial improvement for accurate determination of RR. Also, this performance is comparable to that of the state-of-the-art RR estimation method [8] and the SQF method currently implemented for additional comparison (MAE of 2.6 brpm), despite the limited availability of real world data with annotations for deep learning approach for RR estimation; and the deep learning performance can be drastically improved further with abundant high quality real-world labeled data.

On the other hand, the real datasets employed in this study are mostly, if not all, stationary data. Subsequently, both the present deep learning model and the classic method might not generalize well to the ambulatory PPG data from physical activities. However, this problem can be mitigated by feeding annotated PPG data with motion artifact via online learning or retraining, when such data become available in the future. Furthermore, the synthetic data can include different levels of motion activity to mimic more realistic real-world scenarios by learning the data collected from controlled experiments. Although the deep learning models' training requires sophisticated computational infrastructure, the computational cost for the model's inference is much less, which can be beneficial for embedded applications.

Despite the labeled PPG data for current RR estimation are limited, a vast unlabeled PPG data from the widespread use of PPG sensors in fitness devices and smart watches can be very useful for unsupervised or semi-supervised machine learning approaches to learn the representation and input for deep learning-based RR estimation. Also, the signal quality of the PPG data can be assessed before we feed the PPG data to the deep learning model. This can help us reject some inaccurate RR estimations or provide confidence scores for the RR estimations, and thus can further improve the performance of the model.

In summary, with sufficient real-world data and reference ground truth, the proposed deep learning approach can be valuable for RR or other vital sign monitoring using PPG and other physiological signals.

REFERENCES

[1] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. Pimentel, A. E. Johnson, J. Alastruey, L. Tarassenko, P. J. Watkinson, R. Beale, and D. A. Clifton, "Breathing Rate Estimation from the Electrocardiogram and Photoplethysmogram: A Review," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 2–20, 2018.

[2] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiological Measurement*, vol. 37, no. 4, pp. 610–626, 2016.

[3] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI).* IEEE, 2017, pp. 141–144.

[4] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, "Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks," *Sensors*, vol. 19, no. 14, p. 3079, 2019.

[5] D. Biswas, L. Everson, M. Liu, M. Panwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Acharyya, C. Van Hoof, and M. Konijnenburg, "CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 2, pp. 282–291, 2019.

[6] S. I. Nikolenko, "Synthetic data for deep learning," 2019.

[7] N. Selvaraj, G. Nallathambi, and P. Kettle, "A Novel Synthetic Simulation Platform for Validation of Breathing Rate Measurement," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2018-July, pp. 1177–1180, 2018.

[8] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–1953, 2013.

[9] G. A. L., A. L. A. N., G. Leon, H. J. M., I. P. Ch., M. R. G., M. J. E., M. G. B., P. Chung-Kang, and S. H. Eugene, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, pp. e215–e220, jun 2000.

[10] A. Schäfer and K. W. Kratky, "Estimation of breathing rate from respiratory sinus arrhythmia: comparison of various methods," *Annals of Biomedical Engineering*, vol. 36, no. 3, p. 476, 2008.

[11] A. M. Chan, N. Ferdosi, and R. Narasimhan, "Ambulatory respiratory rate detection using ECG and a triaxial accelerometer," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 79, pp. 4058–4061, 2013.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10 437–10 453, dec 2015.

[13] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951–2959.

[14] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, 2015.