RESEARCH ARTICLE

Using self-supervised feature learning to improve the use of pulse oximeter signals to predict paediatric hospitalization

[version 1; peer review: 2 approved with reservations]

Paul Mwaniki ¹, Timothy Kamanu², Samuel Akech ¹, Dustin Dunsmuir³, J. Mark Ansermino⁴, M.J.C Eijkemans⁵

¹Kenya Medical Research Institute/Wellcome Trust Research Programme, Nairobi, Kenya

²School of Mathematics, University of Nairobi, Nairobi, Kenya

³Digital Health Innovation Lab, BC Children's Hospital Research Institute, Vancouver, British Columbia, Canada

⁴Department of Anesthesiology, Pharmacology & Therapeutics, University of British Columbia, Vancouver, British Columbia, Canada

⁵Julius Centre for Health Sciences and Primary Care, Department of Medical Humanities, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

V1 First published: 28 Sep 2021, 6:248
<https://doi.org/10.12688/wellcomeopenres.17148.1>
 Latest published: 01 Feb 2023, 6:248
<https://doi.org/10.12688/wellcomeopenres.17148.2>

Abstract

Background: The success of many machine learning applications depends on knowledge about the relationship between the input data and the task of interest (output), hindering the application of machine learning to novel tasks. End-to-end deep learning, which does not require intermediate feature engineering, has been recommended to overcome this challenge but end-to-end deep learning models require large labelled training data sets often unavailable in many medical applications. In this study, we trained machine learning models to predict paediatric hospitalization given raw photoplethysmography (PPG) signals obtained from a pulse oximeter. We trained self-supervised learning (SSL) for automatic feature extraction from PPG signals and assessed the utility of SSL in initializing end-to-end deep learning models trained on a small labelled data set with the aim of predicting paediatric hospitalization. **Methods:** We compared logistic regression models fitted using features extracted using SSL with end-to-end deep learning models initialized either randomly or using weights from the SSL model. We also compared the performance of SSL models trained on labelled data alone (n=1,031) with SSL trained using both labelled and unlabelled signals (n=7,578).

Results: The SSL model trained on both labelled and unlabelled PPG signals produced features that were more predictive of hospitalization compared to the SSL model trained on labelled PPG only (AUC of logistic regression model: 0.78 vs 0.74). The end-to-end deep learning model had an AUC of 0.80 when initialized using the SSL model trained on all PPG signals, 0.77 when initialized using SSL trained on labelled data only, and 0.73 when initialized randomly.

Open Peer Review

Approval Status ? ?

	1	2
version 2 (revision) 01 Feb 2023		
version 1 28 Sep 2021	 view	 view
<p>1. Ciira wa Maina, Dedan Kimathi University of Technology, Nyeri, Kenya</p> <p>2. Shaun Davidson , University of Oxford, Oxford, UK</p> <p>Any reports and responses or comments on the article can be found at the end of the article.</p>		

Conclusions: This study shows that SSL can improve the classification of PPG signals by either extracting features required by logistic regression models or initializing end-to-end deep learning models. Furthermore, SSL can leverage larger unlabelled data sets to improve performance of models fitted using small labelled data sets.

Keywords

Signal processing, Self-supervised learning, Photoplethysmography, Deep learning, Hospitalization

This article is included in the [KEMRI | Wellcome Trust](#) gateway.

Corresponding author: Paul Mwaniki (pmwaniki@kemri-wellcome.org)

Author roles: **Mwaniki P:** Conceptualization, Formal Analysis, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kamanu T:** Conceptualization, Writing – Review & Editing; **Akech S:** Conceptualization, Writing – Review & Editing; **Dunsmuir D:** Data Curation, Software, Writing – Review & Editing; **Ansermino JM:** Conceptualization, Writing – Review & Editing; **Eijkemans MJC:** Conceptualization, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported through the DELTAS Africa Initiative [107754/Z/15/Z-DELTAS Africa SSACAB]. P.M also received additional support from DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from Wellcome [107769/Z/10/Z] and the UK government. SA was supported by the Initiative to Develop African Research Leaders (IDeAL) Wellcome award [107769]. All authors acknowledge the support of Wellcome to the Kenya Major Overseas Programme [203077].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Mwaniki P *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Mwaniki P, Kamanu T, Akech S *et al.* **Using self-supervised feature learning to improve the use of pulse oximeter signals to predict paediatric hospitalization [version 1; peer review: 2 approved with reservations]** Wellcome Open Research 2021, 6:248 <https://doi.org/10.12688/wellcomeopenres.17148.1>

First published: 28 Sep 2021, 6:248 <https://doi.org/10.12688/wellcomeopenres.17148.1>

Introduction

Pulse oximeters are used in routine clinical practice to detect hypoxia (low blood oxygen) and measure heart rate (WHO, 2016). Photoplethysmography (PPG) is the underlying mechanism employed by pulse oximeters. The resulting PPG waveform contains additional information on the respiratory and cardiovascular systems that can help diagnose various medical conditions or pathophysiological states.

Analysis of bio-signals such as PPG typically involves signal pre-processing, feature extraction, and classification or regression using the extracted features. Filters are applied to the signal during the pre-processing stage to eliminate noise such as motion artefacts before hand-crafted features required for regression or classification are extracted. Such features are extracted from signals in the original time-domain or after frequency decomposition using Fourier or wavelet transform (Cvetkovic *et al.*, 2008). Hand-crafted features require domain knowledge on the relationship between the waveform and the task of interest, limiting the application of bio-signals to novel tasks.

Deep learning models can bypass signal pre-processing and feature extraction required for classification and regression tasks, speeding up model development in tasks where the relationship between the raw signal and outcome of interest is not well understood in a process referred to as end-to-end deep learning. End-to-end deep learning models have been applied successfully to computer vision (CV), natural language processing (NLP), and raw audio signals (Devlin *et al.*, 2019; He *et al.*, 2015; Schmidhuber, 2015). Automatic feature extraction and classification using end-to-end deep learning can be used to analyse PPG signals, but the models are likely to overfit when trained on small data sets (Bian *et al.*, 2020; Panwar *et al.*, 2020).

Pre-trained models for either feature extraction or model initialization have been shown to improve models trained using small labelled data sets in CV and NLP tasks, but such models are generally unavailable for other tasks (del Rio *et al.*, 2018; Oquab *et al.*, 2014). In this study, we used self-supervised learning (SSL) methods that do not require the availability of pre-trained models. SSL uses a pretext task that learns useful representation (features) from the inputs and a downstream task that solves the problem of interest using the extracted features. Pretext tasks are either generative or discriminative tasks; generative models such as auto-encoders aim to reconstruct the input or generate data from the same distribution as inputs, whilst discriminative approaches such as contrastive learning derive labels from the inputs. An example of contrastive learning in CV is instance classification, where each image in the training data set is treated as a class, and multiple views of each image are obtained through stochastic image augmentation (Chen *et al.*, 2020; He *et al.*, 2020). An encoder model is then trained so that the distances between the embedding of positive pairs (views from the same image) are closer compared to those of negative pairs, which are views from different images.

Self-supervised learning using contrastive methods has been used for extracting features from electroencephalogram (EEG) signals by training an encoder to learn the position of signal segments, with neighbouring segments regarded as positive pairs and those further apart treated as negative pairs (Banville *et al.*, 2019; Banville *et al.*, 2020). However, such an approach limits contrastive learning methods that benefit from having many negative pairs because the length of each signal limits the number of negative pairs. Therefore, we use contrastive learning with a different pretext task where randomly sampled segments from the same patient are treated as positive pairs, and those from different patients are treated as negative pairs.

This study explores the utility of SSL learning in extracting features from raw PPG signals and initializing end-to-end deep learning models to predict hospitalization given raw PPG signals. We use features extracted using contrastive learning to predict hospitalization using logistic regression. We also relate extracted features to known physiological parameters such as respiratory rate, heart rate and oxygen saturation (SpO₂), which are associated with severity of illness using linear regression. We compare end-to-end deep learning models for predicting hospitalization initialized randomly with those initialized using weights of the SSL model. Finally, we assess the utility of an unlabelled PPG signal data set sourced from a different study in improving SSL models. That is, we compared the prediction performance of features extracted using an SSL model trained on labelled data alone with an SSL model trained using both labelled and unlabelled data.

Methods

Study setting and data

Labelled PPG signals. The study was conducted at the paediatric outpatient department of Mbagathi county referral hospital in Nairobi, Kenya, between January and June 2018. The study nurse recruited eligible children aged 2–60 months presenting with medical illnesses before they were reviewed by the attending clinician. The study nurse collected raw PPG signals using a pulse oximeter connected to an android tablet. PPG signals were collected for up to 60 seconds for each patient, and respiratory rate was obtained using an android application that estimated respiratory rates by tapping the tablet's screen with every respiratory cycle (Karlen *et al.*, 2014). The study nurse recorded study participants' demographic information together with clinical signs and symptoms. The facility physicians, who were not part of the study and without access to the study data, decided on whether to admit any patient. The study nurse followed up on patients who were not admitted via telephone calls approximately two weeks after the hospital visit to determine if the patient recovered from the ailment without required admission later or in a different facility. More details about study procedures are described elsewhere (Mawji *et al.*, 2021).

The study collected PPG signals from 1,031 children, of whom 125 (12.2%) were hospitalized. The median age of recruited children was 16 months with an IQR of 9 – 30 months. The PPG signals had median signal quality index (SQI) of 80.5% (IQR 57.25% - 93.0%). Data from 20% of the children were

set aside for final model performance evaluation, while the rest was used for model training and validation.

This study was approved by the Scientific and Ethics Review Unit (SERU) of the Kenya Medical Research Institute (certificate number, SERU/3407). The study nurse obtained written informed consent from the caregiver of recruited children before data collection except in emergency cases where consent was obtained after the child was stable to avoid delaying emergency care.

Unlabelled PPG signals. We obtained additional unlabelled PPG signals from the Smart Discharges (SD) study following an open access request to the Pediatric Sepsis CoLab to evaluate the utility of additional unlabelled data in improving the performance of self-supervised learning models (Wiens *et al.*, 2021a; Wiens *et al.*, 2021b). We compared self-supervised models fitted using labelled data alone with models fitted using both labelled and unlabelled PPG signals. The SD study recruited children below 60 months old, hospitalized in six public hospitals in Uganda with suspected or confirmed infection. The PPG signals were treated as unlabelled despite being collected from hospitalized children because they were not collected at triage or soon after the patient presented at the hospital. The study staff recorded two 60 seconds PPG signals at both admission and discharge. There were 2707 children aged below six months and 3840 aged between 6 and 60 months.

PPG signal acquisition and pre-processing

The pulse oximeters used for PPG signal acquisition for both Kenya and Uganda studies were developed by Lionsgate Technologies Medical (Vancouver, Canada) and made use of photoplethysmographic pulses in the red and infrared wavelengths. The tablet application saved PPG signals that had been downsampled to 80Hz for later analysis.

We split the PPG signals into segments of 10 seconds with two seconds overlap between consecutive segments and normalized each channel using mean and standard deviation calculated from the Mbagathi hospital (labelled) data set.

Feature extraction using contrastive learning

We fitted an encoder to extract features from raw PPG signals using Pytorch 1.7 (Paszke *et al.*, 2019). The architecture for the encoder was similar to ResNet50 with several modifications (Zagoruyko & Komodakis, 2017). We used one-dimensional convolution layers because PPG signals are one dimensional. The first convolutional layer had two channels instead of three because PPG signals have two channels (red and infrared) instead of the red, blue, and green channels in colour images. Max-pooling was applied after every convolutional layer group to increase the network's receptive field and reduce computation cost. The fully connected layer of ResNet50 was replaced with two fully connected layers with 32 units each. Similar to (Chen *et al.*, 2020), the last fully connected layer was discarded after training.

Two contrastive learning models were trained: one model was trained using PPG signals from Mbagathi hospital only (labelled data), and the other was trained using PPG signals from both

Mbagathi hospital and unlabelled PPG signals from the SD study. The two contrastive models were used to extract features for classification using logistic regression and initialize the weights of the end-to-end deep learning models.

Pretext task

The encoder was trained so that embeddings of two segments from the same patient, referred here as positive pairs, were closer than those of two segments from different patients (negative pairs). The unlabelled data from the SD study had multiple PPG recordings for each patient: two admission and two discharge recordings. We treated the admission and discharge recordings as coming from different patients because a substantial amount of time may have passed between the recordings (hospitalization length of stay). A mini-batch would consist of N pairs of segments during training, with each pair sampled from segments from the same patient. Similar to (Chen *et al.*, 2020; He *et al.*, 2020), negative pairs for a given patient were obtained by taking the $2(N-1)$ segments in the mini-batch that did not belong to that patient. We also tried a different pretext task that classified whether two PPG segments are obtained from multiple augmentations of the same PPG segment. This pretext task was not able to learn useful representations of the PPG signals and was therefore dropped.

Noise contrastive estimation (NCE) loss was used to train the encoder (Gutmann & Hyvärinen, 2010). Dot product was used to measure the distance between two embeddings. The contrastive loss for segment j in a mini-batch of N pairs of segments is shown in Equation 1.

$$l_j = -\log \frac{\exp(q \cdot k_+ / \tau)}{\sum_{i=0}^{2(N-1)} \exp(q \cdot k_i / \tau)} \quad (\text{Equation 1})$$

where q is a query embedding, k_+ is the embedding from the positive pair, k_i are the embeddings of $2(N-1)$ negative and one positive segments, and τ is a temperature hyper-parameter.

Data augmentation was applied to PPG signals in the training data set as part of the model training pipeline. Data augmentation was done by adding Gaussian noise to the signals and window slicing and permutation (Um *et al.*, 2017). Gaussian noise with standard deviation ranging between $1e-5$ and $1e-2$ was added to a proportion of training signals. Window slicing and permutation was done by varying the number of windows and the proportion of augmented signals.

Hyper-parameter tuning was done using Asynchronous Successive Halving Algorithm (ASHA) using the ray-tune library in Python (Li *et al.*, 2018; Liaw *et al.*, 2018). The tuned hyper-parameters included regularization parameters (dropout and weight decay), data augmentation (proportion of signals with augmentation and the number of slices), batch size, and temperature parameter of NCE loss.

Extracted features were evaluated qualitatively using t-distributed stochastic network embeddings (t-SNE). Three scatter plots of the first and second components of t-SNE were shaded using values of heart rate, respiratory rate and SpO_2 . Features for each patient were obtained by averaging embedding of all segments belonging to that patient.

Classification and regression using features extracted using contrastive learning

Features extracted using contrastive learning were used as predictors of hospitalization with logistic regression models. The loss of the logistic regression model was weighted by the inverse of class frequency to address class imbalances in the outcome. Consequently, children with predicted probability above 0.5 were presumed to be hospitalized. Linear regression was used to predict heart rate, respiratory rate and SpO₂. We conducted univariate feature selection for both linear and logistic regression by selecting features that had high mutual information with the outcome. Grid search hyper-parameter optimization was used to select optimal hyper-parameters for the number of features to keep and weight decay. Bagging was applied to both logistic and linear regression models to reduce variance in predictions. Model performance was assessed using the area under the curve (AUC) for classification models, and root mean square error (RMSE) and coefficient of determination (R²) for regression models.

End-to-end classification using deep learning

The end-to-end deep learning models for predicting hospitalization were identical to the encoder used for contrastive learning, except that the last fully connected layer had a single output. We compared end-to-end deep learning models initialized randomly with those initialized using weights learned by contrastive learning models. The models were trained on soft labels (label smoothing) recommended for noisy targets, and

the minority class was oversampled to address class imbalance (Müller *et al.*, 2019; Szegedy *et al.*, 2015). We tuned hyper-parameters for regularization (dropout, weight decay, data augmentation), batch size, learning rate, and label smoothing parameter using the ASHA algorithm. Python code for the analysis is available on [Github](https://github.com/pmwaniki/ppg-analysis/releases/tag/v1.0.0) (<https://github.com/pmwaniki/ppg-analysis/releases/tag/v1.0.0>).

Results

The contrastive learning model trained on both labelled and unlabelled PPG signals had higher accuracy in classifying whether two PPG segments were obtained from the same patient than the model trained on labelled data only (validation accuracy, 91% vs 73%). Qualitative evaluation of extracted features using t-SNE showed that the features were correlated with heart rate, respiratory rate, and SpO₂ (Figure 1). Logistic regression models for predicting hospitalization using the extracted features had AUC of 0.74 (sensitivity 0.62 & specificity 0.64) for the contrastive learning model trained on labelled data only and 0.78 (sensitivity 0.77 & specificity 0.68) for the contrastive model trained on both labelled and unlabelled data. Features extracted using the contrastive model trained on both labelled and unlabelled data were also more predictive of heart rate (R² 0.82 vs 0.71), respiratory rate (R² 0.35 vs 0.25) and SpO₂ (R² 0.74 vs 0.04). However, the regression models under-estimated respiratory rate measurements above 70 breaths per minute and overestimated SpO₂ measurements below 80 (Figure 2).

Figure 1. t-SNE visualizations of features extracted using SSL shaded by heart rate, respiratory rate and SpO₂. (a) Features are extracted using the contrastive learning model fitted using labelled PPG signals only. (b) Features are extracted using the contrastive learning model fitted using both labelled and unlabelled PPG signals.

Figure 2. Performance of models predicting heart rate, respiratory rate and SpO_2 using features extracted using SSL. (a) Features are extracted using contrastive model fitted using labelled PPG signals only. (b) Features are extracted using contrastive learning fitted using both labelled and unlabelled PPG signals. The red dotted line shows test data points where the predicted value equals the observed value (perfect prediction).

End-to-end deep learning models for predicting hospitalization had an AUC of 0.73 when initialized randomly. The AUC of end-to-end models rose to 0.77 when initialized using the SSL model trained on labelled data only, and 0.80 for model initialized using the SSL model trained on both labelled and unlabelled data (Table 1).

Discussion

Key findings

This study sought to evaluate the utility of SSL in extracting features from raw PPG signals and initializing end-to-end deep learning models with the aim of predicting hospitalization.

Moreover, a comparison is made between SSL models trained using labelled data only with models trained with additional unlabelled data. Self-supervised models performed better when trained on labelled and unlabelled data despite the unlabelled data set originating from a different setting. Features extracted using SSL model trained using labelled and unlabelled data were more predictive of hospitalization, heart rate, respiratory rate, and SpO_2 than features extracted using the model trained on labelled data only. Weights from SSL models trained on both labelled and unlabelled data were better at initializing end-to-end models. Weights from the self-supervised model trained on labelled data alone were better at initializing

Table 1. Performance of classification models for predicting hospitalization. SSL models are used to extract features for classification using logistic regression models and initializing end-to-end deep learning models.

Model	Initialization	Precision	Sensitivity	Specificity	AUC
Contrastive learning & Logistic Regression	Contrastive: Labelled & Unlabelled	0.26	0.77	0.68	0.78
	Contrastive: Labelled	0.2	0.62	0.64	0.74
End to end	Contrastive: Labelled & Unlabelled	0.27	0.85	0.67	0.8
	Contrastive: Labelled	0.21	0.81	0.56	0.77
	Random	0.2	0.73	0.58	0.73

end-to-end models than random initialization, suggesting that self-supervised models may still be useful when all data is labelled.

We trained a contrastive model such that embeddings of segments from the same patient were closer than embedding from different patients. The pretext task may be well suited for extracting features from PPG and other bio-signals because the underlying physiological parameters are unlikely to change significantly within a short time. The SSL model trained to embed two augmentations of the same PPG segment did not extract useful features despite the pretext task being successful for image tasks, likely due to insufficient data augmentation procedures. Successful augmentation of bio-signals may require introducing noise similar to what is encountered in real life, such as motion artefacts for PPG signals. Audio tasks have benefited from data augmentation techniques such as time stretching and pitch shift (Wei *et al.*, 2020). However, such augmentation may be unsuitable for bio-signals. Time stretching, for instance, would alter the underlying measurement of heart rate, which may negatively affect performance on downstream tasks.

Our study is in agreement with previous studies that showed that PPG signals are predictive of hospitalization. (Garde *et al.*, 2016) achieved an AUC of 0.75 in predicting hospitalization using logistic regression based on age, gender, respiratory rate, and hand-crafted features derived from PPG signals using traditional signal processing techniques. Features were extracted from the PPG signal in the time domain based on pulse-to-pulse interval (PPI) statistics, frequency domain after computing the power spectrum density, and additional features computed from SpO₂. We used SSL for automatic feature extraction eliminating the need for extensive feature engineering or intermediate calculations of SpO₂. (Mawji *et al.*, 2021) achieved an AUC of 0.88 on the same data set using logistic regression and clinical signs and symptoms as predictors. Although clinical signs and symptoms are more predictive of hospitalization than raw PPG signals, clinical features are subjective and require more effort to collect. Furthermore, SSL learning can provide additional features from the bio-signals to use alongside clinical signs and symptoms to produce more predictive models.

Implications

Features extracted using self-supervised learning were useful for predicting heart rate, respiratory rate, and SpO₂ despite the model not being optimized to solve those tasks, implying that the features may contain information on yet unknown physiological parameters contained in the PPG signal. Such parameters can enhance the development of novel applications using bio-signals. Such applications may include non-invasive continuous disease monitoring using pulse oximeters, improved triaging of patients seeking care and the development of cheaper more objective diagnostic tools.

Limitations

We used SpO₂ and heart rate measures obtained using a pulse oximeter while the respiratory rate was approximated after tapping on the respiratory rate application with each respiratory cycle. This respiratory rate approximation is a single value (not continuously monitored) and likely to contain noise and this may underestimate the predictive performance of extracted features.

Although we used the raw wave-form, the pulse oximeters were from a single manufacturer. Therefore, models trained in this study are unlikely to perform well when applied to devices from different manufacturers without retraining the models. Furthermore, device manufacturers may apply signal pre-processing procedures which are proprietary and implemented in closed source software, making standardization of model development unfeasible. Moreover, it is unclear whether the utility of incorporating unlabelled data using self-supervised learning would remain if the labelled and unlabelled data were collected using pulse oximeters from different manufacturers.

The deep learning models used for both SSL and end-to-end deep learning are computationally intensive compared to traditional signal processing techniques. Consequently, we were not able to estimate the variability in model performance using K-fold cross-validation. However, we used the save test data set to compare all models and SSL improved performance of all models suggesting the observed benefit is unlikely to be due to chance.

Conclusion

In conclusion, we demonstrated that SSL using contrastive learning is useful for extracting features from raw PPG signals and initializing end-to-end deep learning models. SSL could speed up the development of novel applications using bio-signals such as PPG by not requiring extensive domain knowledge about the relationship between the waveform and the task of interest. SSL can also leverage large unlabelled data sets to improve the predictive performance of models fitted using small labelled data sets.

Data availability

Underlying data

Harvard Dataverse: Replication Data for: Assessment of Automation of Paediatric Triage using Pulse Oximetry in a Kenyan Public Hospital, <https://doi.org/10.7910/DVN/KQ4DNK> (Mwaniki *et al*, 2021).

This project contains the following underlying data:

- Data.csv (Demographic, clinical, and follow-up data of recruited participants)
- PaediatricTriage_DataDictionary.tab (Data dictionary for Data.csv)
- pulse_oximeter_data.zip (zipfile containing the red and infrared pulse oximeter waveforms for each patient)

The data is under restricted access because it contains potentially identifying information on recruited patients. External data sharing guidelines are available through the [KEMRI Data Governance page](#). Data sharing requests should be addressed to dgc@kemri-wellcome.org.

Scholars Portal Dataverse: 6-60m Observation - Pulse Oximetry (dataset) ~ Smart Discharges, <https://doi.org/10.5683/SP2/PHX4C5> (Wiens *et al*, 2021a)

Scholars Portal Dataverse: <6m Observation - Pulse Oximetry (dataset) ~ Smart Discharges, <https://doi.org/10.5683/SP2/ZDDFZL> (Wiens *et al*, 2021b).

The Scholars Portal Dataverse projects are under restricted access. Data requests should be made to the [Pediatric Sepsis Data Colab](#).

Extended data

Analysis code available from: <https://github.com/pmwaniki/ppg-analysis/>

Archived analysis code as at time of publication: <https://doi.org/10.5281/zenodo.5501144> (Mwaniki, 2021).

License: MIT license

Acknowledgements

The authors would like to thank the administration and staff at the outpatient pediatric department of Mbagathi County Hospital, and the caregivers/parents of recruited participants. We acknowledge Purity Makena, the study nurse for data collection and consenting all study participants. We would also like to thank Matthew O. Wiens and the Pediatric Sepsis CoLab for providing additional PPG signals from the Smart Discharges study. This work is published with the permission of the Director of KEMRI.

References

- Banville H, Albuquerque I, Hyvärinen A, *et al.*: **Self-supervised representation learning from electroencephalography signals**. *ArXiv: 1911.05419 [Cs, Eess, Stat]*. 2019.
[Reference Source](#)
- Banville H, Chehab O, Hyvärinen A, *et al.*: **Uncovering the structure of clinical EEG signals with self-supervised learning**. *ArXiv: 2007.16104 [Cs, Eess, q-Bio, Stat]*. 2020.
[Reference Source](#)
- Bian D, Mehta P, Selvaraj N: **Respiratory Rate Estimation using PPG: A Deep Learning Approach**. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020; 2020: 5948–5952.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chen T, Kornblith S, Norouzi M, *et al.*: **A Simple Framework for Contrastive Learning of Visual Representations**. *ArXiv: 2002.05709 [Cs, Stat]*. 2020.
[Reference Source](#)
- Cvetkovic D, Übeyli ED, Cosic I: **Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study**. *Digit Signal Process*. 2008; 18(5): 861–874.
[Publisher Full Text](#)
- del Rio F, Messina P, Dominguez V, *et al.*: **Do Better ImageNet Models Transfer Better... For Image Recommendation?** *ArXiv: 1807.09870 [Cs]*. 2018.
[Reference Source](#)
- Devlin J, Chang MW, Lee K, *et al.*: **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. *ArXiv: 1810.04805 [Cs]*. 2019.
[Reference Source](#)
- Garde A, Zhou G, Raihana S, *et al.*: **Respiratory rate and pulse oximetry derived information as predictors of hospital admission in young children in Bangladesh: A prospective observational study**. *BMJ Open*. 2016; 6(8): e011094.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gutmann M, Hyvärinen A: **Noise-contrastive estimation: A new estimation principle for unnormalized statistical models**. In Y. W. Teh & M. Titterton (Eds.), *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings. 2010; 9: 297–304.
[Reference Source](#)
- He K, Fan H, Wu Y, *et al.*: **Momentum Contrast for Unsupervised Visual Representation Learning**. *ArXiv: 1911.05722 [Cs]*. 2020.
[Reference Source](#)
- He K, Zhang X, Ren S, *et al.*: **Deep Residual Learning for Image Recognition**. *CoRR, abs/1512.03385*. 2015.
[Reference Source](#)
- Karlen W, Gan H, Chiu M, *et al.*: **Improving the Accuracy and Efficiency of Respiratory Rate Measurements in Children Using Mobile Devices**. *PLoS One*. 2014; 9(6): e99266.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li L, Jamieson K, Rostamizadeh A, *et al.*: **Massively Parallel Hyperparameter Tuning**. 2018.
[Reference Source](#)
- Liaw R, Liang E, Nishihara R, *et al.*: **Tune: A Research Platform for Distributed**

Model Selection and Training. *ArXiv: 1807.05118 [Cs, Stat]*. 2018.

[Reference Source](#)

Mawji A, Akech S, Mwaniki P, *et al.*: **Derivation and internal validation of a data-driven prediction model to guide frontline health workers in triaging children under-five in Nairobi, Kenya** [version 3; peer review: 2 approved]. *Wellcome Open Res.* 2021; **4**: 121.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Müller R, Kornblith S, Hinton G: **When Does Label Smoothing Help?** 2019.

[Reference Source](#)

Mwaniki P, Dunsmuir D, Ansermino MJ, *et al.*: **Replication Data for: Assessment of Automation of Paediatric Triage using Pulse Oximetry in a Kenyan Public Hospital.** Harvard Dataverse, V1, 2021. UNF:6: 1vb/pS9mYzAo+7AxBMYSdQ== [fileUNF]. <http://www.doi.org/10.7910/DVN/KQ4DNK>

Mwaniki P: **pmwaniki/ppg-analysis: 1.0.0 (v1.0.0).** *Zenodo*. 2021.

<http://www.doi.org/10.5281/zenodo.5501144>

Oquab M, Bottou L, Laptev I, *et al.*: **Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks.** 2014 *IEEE Conference on Computer Vision and Pattern Recognition*. 2014; 1717–1724.

[Publisher Full Text](#)

Panwar M, Gautam A, Biswas D, *et al.*: **PP-Net: A Deep Learning Framework for PPG-Based Blood Pressure and Heart Rate Estimation.** *IEEE Sensors Journal*. 2020; **20**(17): 10000–10011.

[Publisher Full Text](#)

Paszke A, Gross S, Massa F, *et al.*: **PyTorch: An Imperative Style, High-Performance Deep Learning Library.** In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* 32. Curran Associates, Inc. 2019; 8024–8035.

[Reference Source](#)

Schmidhuber J: **Deep learning in neural networks: An overview.** *Neural Netw.* 2015; **61**: 85–117.

[PubMed Abstract](#) | [Publisher Full Text](#)

Szegedy C, Vanhoucke V, Ioffe S, *et al.*: **Rethinking the Inception Architecture for Computer Vision.** 2015.

[Reference Source](#)

Um TT, Pfister FMJ, Pichler D, *et al.*: **Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks.** *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 2017; 216–220.

[Publisher Full Text](#)

Wei S, Zou S, Liao F, *et al.*: **A Comparison on Data Augmentation Methods Based on Deep Learning for Audio Classification.** *J Phys Conf Ser.* 2020; **1453**: 012085.

[Publisher Full Text](#)

WHO: **Oxygen therapy for children: A manual for health workers.** *World Health Organization*. 2016.

[Reference Source](#)

Wiens M, Kabakyenga J, Kumbakumba E, *et al.*: **6-60m Observation—Pulse Oximetry (dataset) Smart Discharge.** (Version V1). *Scholars Portal Dataverse*. 2021a.

<http://www.doi.org/10.5683/SP2/PHX4C5>

Wiens M, Kabakyenga J, Kumbakumba E, *et al.*: **<6m Observation - Pulse Oximetry (dataset) ~ Smart Discharges.** *Scholars Portal Dataverse*, V1. 2021b. <http://www.doi.org/10.5683/SP2/ZDDFZL>

Zagoruyko S, Komodakis N: **Wide Residual Networks.** *ArXiv: 1605.07146 [Cs]*. 2017.

[Reference Source](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 25 July 2022

<https://doi.org/10.21956/wellcomeopenres.18941.r49279>

© 2022 Davidson S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Shaun Davidson

Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK

This paper explores using self-supervised learning to automatically detect PPG features best suited to predicting paediatric hospitalisation from a set of 60-second recordings. These features are then fed into both a logistic regression and deep learning model, and performance is assessed. The features derived via self-supervised learning were found to improve performance compared to random model initialisation. The paper is well-written and methodologically interesting. Additionally, the authors are to be commended for making the Python code of their methodology available to readers. However, this reviewer feels that there are several improvements to the manuscript that should be made.

1. The authors note that "*self-supervised models performed better when trained on labelled and unlabelled data...*" (Discussion, paragraph 1). This reviewer feels that this statement is too broad and needs to be framed in the context of the relatively small size of the labelled data set (1031 recordings) and the larger size of the unlabelled dataset (approx. 13000 recordings). Clearly, additional unlabelled data can improve self-supervised learning in cases where only a small labelled dataset is available, and, as such cases are common, this result has clear and important benefits. However, the statement as written could be read as suggesting that adding unlabelled data, regardless of the size of your labelled or unlabelled dataset, improves the results of self-supervised learning, which this reviewer does not feel is supported by the results in the paper and is unlikely to be the case (a sufficiently large labelled dataset would likely perform identically or better, but require significant more manual annotation).
2. This reviewer feels (in line with reviewer 1) that the comparison between a model initialised using self-supervised learning compared to random initialisation is not quite a 'like-for-like' comparison. Most PPG algorithms, as stated in the manuscript's introduction, tend to use hand-crafted features derived from decades of signal processing and medical work. Surely it would be worth comparing the results of the self-supervised learning methods with these methods which are more representative of the existing 'state-of-the-art' and involve an

equivalent, if different and manual, signal pre-processing layer.

3. This reviewer feels that the novelty and clinical relevance of this work need to be made more clear. While self-supervised learning can be useful in cases where "*domain knowledge on the relationship between the waveform and the task of interest*" is not available, PPG is definitely not one of these cases (as mentioned in the previous point). Additionally, despite all of this existing domain knowledge and the general availability of PPG signals, this reviewer is not aware of any clinical desire or inclination to move towards using PPG as a sole means of triage (as opposed to using PPG-derived HR and SpO2 as part of a broader framework of vital signs for triage). How would the authors expect this approach to be received by clinicians or employed in the clinic, and what benefits do they feel it could provide?

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Circadian rhythms, wearable technology, non-contact vital sign monitoring

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 02 November 2021

<https://doi.org/10.21956/wellcomeopenres.18941.r46193>

© 2021 Maina C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ciira wa Maina

Department of Electrical and Electronics Engineering, Dedan Kimathi University of Technology, Nyeri, Kenya

The paper presents a model to predict paediatric hospitalisation using the raw photoplethysmography signal obtained from a pulse oximeter. This model relies on self supervised learning to initialise deep neural networks to overcome the difficulty of training deep learning models from scratch using limited data.

Comments

1. The justification for using the ResNet 50 model for the encoder is not clear. Given the limited data, it is possible that a simpler model could lead to better performance.
2. The tSNE visualisations are confusing. What clustering is expected here? Perhaps moving to 3D would improve the visualisation.
3. This paper should compare the predictions based on the neural network with a model that extracts parameters of interest such as heart rate, respiratory rate and oxygen saturation from the raw photoplethysmography signal and then uses these to predict paediatric hospitalisation.
4. The authors should comment on how this model would be deployed and used in the clinic.
5. In the comparison between the predicted values of heart rate and the labelled values (Figure 2) the authors should compare the values obtained to other methods in the literature. In addition, what influence does the anatomical location from which the signal is obtained have on the accuracy?

Minor Comments

1. In describing the Gaussian noise used to corrupt the signal, signal-to-noise ratio is more useful than the variance in absolute terms.
2. In the description of the terms in Equation 1, the 'i' in 'ki' should be a subscript.
3. On the last paragraph of page 7, edit the line "However, we used the save test data set to compare..."
4. In the conclusion, include a space in "... models.SSL..."

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Machine learning, digital signal processing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.