2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Transfer learning of CNN-based signal quality assessment from clinical to non-clinical PPG signals

Serena Zanelli[1], Mounim A. El Yacoubi[2], Magid Hallab[3], Mehdi Ammi[4]

Abstract— **Photoplethysmography (PPG) is a non-invasive and cost-efficient optical technique used to assess blood volume variation inside the micro-circulation. PPG technology is widely used in a variety of clinical and non-clinical devices in order to investigate the cardiovascular system. One example of clinical PPG device is the pulse oxymeter, while non-clinical PPG devices include smartphones and smartwatches. Such a wide diffusion of PPG devices generates plenty of different PPG signals that differ from each other. In fact, intrinsic device characteristics strongly influence PPG waveform. In this paper we investigate transfer learning approaches on a Covolutional Neural Network based quality assessment method in order to generalize our model across different PPG devices. Our results show that our model is able to classify accurately signal quality over different PPG datasets while requiring a small amount of data for fine-tuning.**

Clinical relevance— **A precise detection and extraction of high quality PPG segments could enhance significantly the reliability of the medical analysis based on the signal.**

I. INTRODUCTION

Cardiovascular diseases (CVDs) are a group of disorders concerning the heart and blood vessels. CVDs include heart attack, cognitive impairment, heart and kidney failure and peripheral arterial disease. In 2017, CVDs were the first cause of mortality worldwide representing 33% of all global deaths [6]. Electrocardiogram (ECG) and cardiac ecography are considered the gold standard to detect CVDs. However, even if these technologies are becoming more accessible, they still require specialised healthcare personnel in order to be analysed. Photopletysmography (PPG) is an interesting alternative to CVDs gold standard detection methods. PPG is an optical measurement of blood volume changes in the vessels. It is attracting interest from industries and the scientific community thanks to its cost-effective technology and usability. Thanks to its richness in cardiovascular information, researchers have been capable to use the PPG signal to estimate blood pressure [9], arterial stiffness [11] and detect atrial fibrillation (AF) [17]. However, PPG is often corrupted by motion artifacts and noise due to external factors as external temperature and light. In order to conduct a reliable analysis, it is important to detect whether or not the signal is exploitable. Hence, signal quality detection is the first fundamental step to obtain robust PPG analysis. Several traditional methods have been implemented in order to remove low frequency components [16], high frequency noises [7] or motion artifacts [18]. One of the most known traditional signal quality index (SQI) methods is proposed by Elgendi [2] where several indexes have been evaluated based on their capability to classify high and low quality PPG signal. However, sometimes, these methods cannot be applied in real-time as they are composed of several complex steps in order to achieve the desired result [13], [8]. Recently, several machine learning [14] and particularly deep learning methods [10], that offer the possibility to take as input directly the raw signals, have been applied to PPG signal quality analysis. Those methods have obtained better results compared to traditional methods. These approaches, however, are limited to a single device as they do not consider signal device-related variability. PPG can actually be collected using plenty of different technologies and devices. It is possible to identify three main categories of factors that influence PPG waveform. The first one is composed by technology-related factors and it includes the sampling frequency, sensor type, light wavelength, light intensity, filtering and all the other factors linked to the acquisition and pre-processing chain. Secondly we have the subject-related factors. This category includes all the factors strictly related to the subject such as skin colour and temperature, life-style, presence of pathological condition and device position during the measurement. The last category may consist of environmental factors such as external light, humidity and temperature. In addition, the length of the analysed PPG segment in the related studies is considerably longer then one single pulse. Thus, they require a long signal acquisition in order to be applied and they do not provide a single-pulse precision.

In this paper, we investigate transfer learning of a Convolutional Neural Network (CNN) based quality assessment model across different acquisition devices assessing the impact of technology-related factors over the PPG signal. Our contribution is twofold:

- First, before considering the transfer learning mechanism, we propose a baseline CNN-based quality assessment model that takes as input a PPG segment that is much shorter when compared to the cited studies. Since our objective is to classify whether or not a single PPG wave is compromised, the segment length has to contain at least one PPG wave. Thus, we chose 1-second as segment length.

[1]S. Zanelli is with Université Paris 8, Saint-Denis, France zanelli@math.univ-paris13.fr
[2]M. A. El Yacoubi is with Samovar, CNRS, Télécom SudParis, Institut Polytechnique de Paris, Evry, France mounim.el_yacoubi@telecom-sudparis.eu
[3]M. Hallab is with Clinique Bizet, Paris, France magid.axelife@.fr
[4]M. Ammi is with Université Paris 8, Saint-Denis, France mehdi.ammi@univ-paris8.fr

The chosen segment length makes the model capable of operating with a variety of signal lengths without requiring long time PPG acquisitions.

- Secondly, we utilize transfer learning from a CNN-based quality assessment model trained on a finger PPG dataset and fine tune it over two smaller databases, composed by smartphone and bracelet PPG signals. Applying transfer learning, we demonstrate how effectively our model is capable of effectively classify PPG signals collected with different devices.

The rest of the paper is organized as follows. In Section 2, we briefly outline the background of PPG, then we present the used databases and the model architecture. In Section 3, the results are presented. Finally, in Section 4, we conclude and discuss future work.

II. METHODOLOGY

In this section, we present how our proposed approach perform quality assessment of PPG signal. We first introduce the PPG technology, then we describe the different datasets we used and the model architecture.

A. Photoplethysmography

A photoplethysmograph is composed by a Light Emitting Diode (LED) and a photodetector (PD). The light emitted from the diode is reflected or transmitted through the tissues, the arteries and arterioles. The reflected/transmitted light is then detected by the PD. A strong influence on the PPG pulse waveform is given by the anatomical measurement site. In Figure 1, three signals acquired with three different devices are shown. In each signal, both high and low quality PPG signals are represented. The first device is pOpmètre, a finger PPG device from Axelife® that uses a red LED with a sampling frequency of 1KHz [4]. The second device is a iPhone SE, the PPG is extracted from the camera and it is sampled at 120Hz [3]. E4 is the third device and it's a wrist PPG device that uses red and green LED with a sampling frequency of 64 Hz [5].

B. Datasets

Three different datasets have been used in this study. The first dataset (DB1) is composed of finger PPG signals acquired with the pOpmètre device. A total of nine hundred PPG signals of different lengths have been used to train our CNN model. Signals were acquired on five hundred healthy subjects with different ages (range 20 - 80 years old) and different heart rate (71 bpm on average). All the signals were collected in a clinical environment. The second database (DB2) is composed by finger PPG signals collected from a iPhone SE camera. Twenty subjects were asked to collect one signal while they were seated calmly and one while they were moving. The third database (DB3) is composed by wrist PPG signals collected with the E4. Thirteen subjects were asked to collect one signal while they were seated calmly and one while they were moving. DB1 has been used to train our baseline CNN model. Once we trained and optimised the model, we applied transfer learning to the other datasets in order to find the optimal trade-off between quantity of training data and model accuracy.

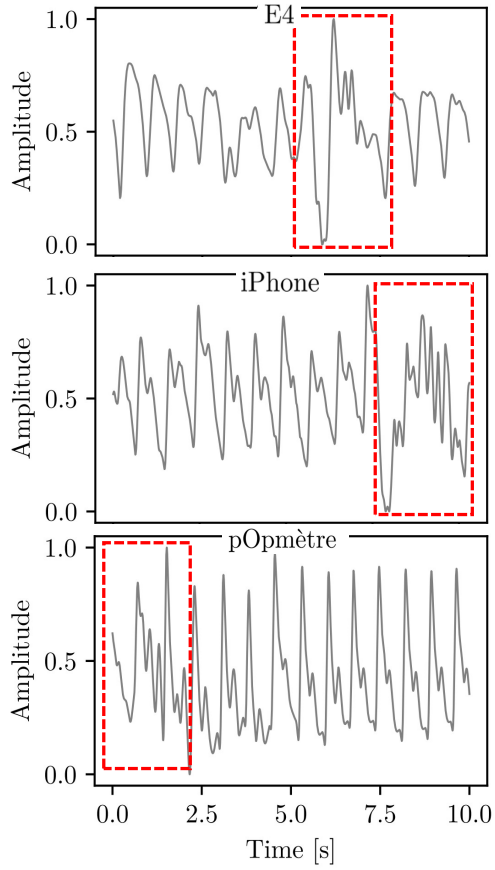

Fig. 1. PPG signal acquired with different devices. In order from the bottom: finger PPG, smartphone camera PPG, wrist PPG. The low quality segments are highlighted with a red dotted line.

TABLE I
DATABASES DECRIPTION

Name	Device	Subjects	Samples	Class 0
DB1	pOpmètre	500	451476	50%
DB2	iPhone	20	7306	48%
DB3	E4	13	7311	49%

C. Pre-processing and annotations

The devices we used have different sampling frequencies. Thus, we first have computed a light filtering in order to remove the high frequency noise and then we have up-sampled/down-sampled all the signals to 100Hz. If one signal contained both high quality and low quality PPG waves, it was rejected. Once all the signals have been visually checked and annotated by a PPG expert, they have been cutted in multiple segments of 100 samples (corresponding to a time length of 1 second) with an overlapping of 95%. Thus, if a signal has been considered as high quality, it was assigned

to class 1. On the contrary, it was assigned to class 0. After segmentation, each sample has been normalised in amplitude between 0 and 1. Table I summarizes the details of the used databases and the percentage of low and high quality PPG segments contained in each DBs.

D. CNN model architecture

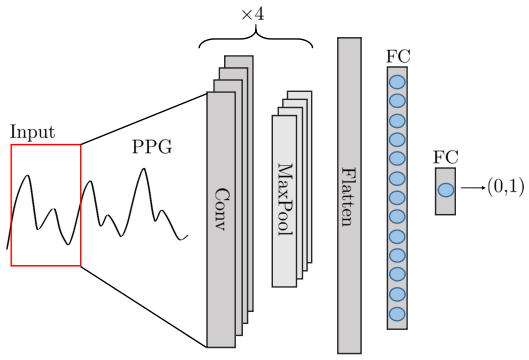

Fig. 2. CNN model architecture.

For this work, we chose the CNN thanks to its capability to auto-extract feature from the signals. The architecture of our CNN model is shown in Figure 2 and it consists of four convolution layers (Conv), followed by a fully connected (FC) layer. Each convolution layer is followed by a maxpooling layer. The output of the max-pooling layer is a feature map containing the most prominent features of the previous feature map. A flattening layer is added after the last convolution layer in order to reduce the dimensions and pass the feature maps to the fully connected layer. Relu activation functions inside the convolution layers have been used, while a sigmoid activation function has been used for the classification layer. The network parameters are optimized with the Adam method, and binary cross-entropy is used as loss function in the output layer. Hyper parameters optimisation has been performed using Autonomio Talos [15]. The number of parameters of the resulting model is 41.000. Taking in consideration a double precision, the model occupies around 2MB of space. The selected hyper parameters are: number of neurons of the dense layer equal to 30, number of filters in the convolution layers equal to 32, dimension of the kernel equal to 11, pool size in the maxpooling layer equal to 2 and batch size of 16. The model takes as input a 1-second signal and classifies it as high quality or bad quality.

III. RESULTS AND DISCUSSION

In this section, we evaluate the results obtained by our network on DB1 and then we investigate the effect of the transfer learning technique on DB2 and DB3.

Our baseline model (trained with DB1) reached a test accuracy of 99.8% with both specificity and sensitivity equal to 99.8%. It scored an accuracy of 86% when tested over

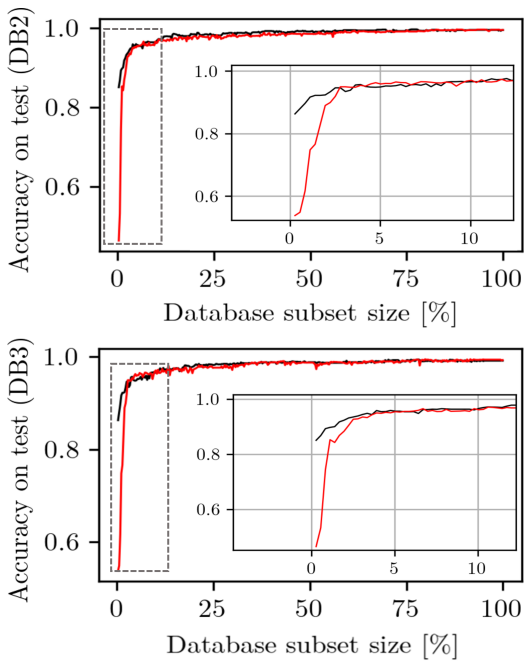

Fig. 3. Upper image: accuracy curves for CNN model trained only with DB2 -red line- and CNN model trained with DB1 and transfer learning over DB2 -black line-. Lower image: accuracy curves for CNN model trained only with DB3 -red line- and CNN model trained with DB1 and transfer learning over DB3 -black line-.

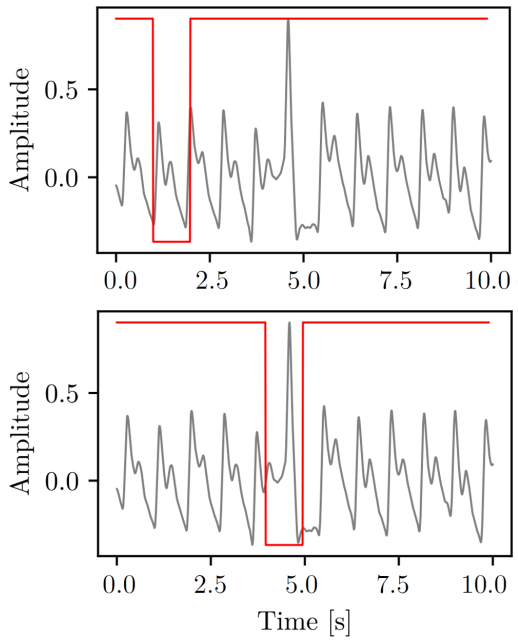

Fig. 4. Example of application of the model without (upper figure) and with (bottom figure) transfer learning over the same smartphone PPG signal. The model classification is plotted in red. High logical values represent high quality PPG while low logical values represent low quality PPG.

DB2. To compare the transfer learning effect, we trained another model with the same architecture and hyper-parameters using only DB2 (no pre-trained weights). In order to find the optimal quantity of data needed to obtain a satisfying accuracy, the two models have been trained and fine tuned with increasing DB2 subset size. Each time the DB2 subset size has been incremented by 10 PPG segments. The same procedure has been conducted over DB3. The test accuracy obtained from the baseline model over DB3 is equal to 81%. As shown in Figure 3, transfer learning allows to reach a satisfying accuracy with a very small dataset. Starting from 4% of the dataset, the results of the two models are comparable. To better understand the results obtained with transfer learning, the same signal has been analysed with the models obtained using the 2.25% of DB2 train set. The model trained only DB2 data scored an accuracy of 86% while the baseline model fine tuned with pre-trained weights scored an accuracy of 93%. In Figure 4, the different classification obtained with and without transfer learning is shown. The 2.25% of the training set is equivalent to a dataset composed of only 81 PPG segments. In other words, it means that it is sufficient to collect 1 minute of PPG signal and fine tune the pre-trained model to obtain an accuracy of 93%. Since the amount of required data is considerably low, the model could be easily fine tuned for each user in order to obtain an accurate personalised classification.

IV. Conclusions

The objective of this work was to investigate transfer learning approaches on a CNN model for quality assessment of PPG signals in order to assess its adaptability to different datasets associated with different acquisition devices. Our model demonstrated great generalization over different datasets when the transfer learning is performed and the transferred model is fine tuned on a small data subset collected from the targeted non-clinical device. Potentially, the model could be shaped, using very few data, over every single user, thus overcoming device and user differences. In addition, the transfer learning technique is well known to speed up the training process, thus reducing the computational cost of the training phase [12].

Future improvements will focus on enhancing the model usability with pathological signals such as the presence of arrhythmia or diabetes. Consequently, segment dimension has to be redesigned in order to take into consideration a larger range of heart rates. In this study, DB1 has been taken as reference but, since results with the other datasets are promising, in the next studies, DB2 and DB3 can be taken as reference. Other transfer learning approaches will be further explored [1].

References

[1] Maxime de Bois, Mounîm A. El Yacoubi, and Mehdi Ammi. Enhancing the interpretability of deep models in heathcare through attention: Application to glucose forecasting for diabetic people. *arXiv*, 2020.

[2] Mohamed Elgendi. Optimal Signal Quality Index for Photoplethysmogram Signals. *Bioengineering*, c:1–15, 2016.

[3] Online: https://www.apple.com/fr/iphone se.

[4] Online: https://www.axelife.fr/.

[5] Online: https://www.empatica.com/research/e4/.

[6] Online: https://www.who.int/gho/publications.

[7] Greeshma Joseph, Almaria Joseph, Geevarghese Titus, Rintu Mariya Thomas, and Dency Jose. Wavelet De-Noising. *2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD)*, pages 1–5, 2014.

[8] Q. Li and G. D. Clifford. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological Measurement*, 33(9):1491–1501, 2012.

[9] Franck Mouney, Teodor Tiplica, Jean-Baptiste Fasquel, Magid Hallab, and Mickael Dinomais. A new blood pressure estimation approach using ppg sensors: Subject specific evaluation over a long-term period. *International Summit Smart City 360°*, pages 45–63, 2021.

[10] Emad Kasaeyan Naeini. A Real-time PPG Quality Assessment Approach for for Healthcare. *Procedia Computer Science*, 151(2018):551–558, 2019.

[11] H. Obeid, P. Boutouyrie, M. Hallab, H. Khettab, and S. Laurent. Calculation of central blood pressure by analyzing the contour of the photoplethysmographic pulse measured at the finger with the popmètre® device. *Archives of Cardiovascular Diseases Supplements*, 10(1):104, 2018.

[12] David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. Carbon emissions and large neural network training, 2021.

[13] Fulai Peng, Zhengbo Zhang, Xiaoming Gou, Hongyun Liu, and Weidong Wang. Motion artifact removal from photoplethysmographic signals by combining temporally constrained independent component analysis and adaptive filter. *Biomedical engineering online*, 13(1):1–14, 2014.

[14] Nikhilesh Pradhan, Sreeraman Rajan, and Andy Adler. Evaluation of the signal quality of wrist-based photoplethysmography. *Physiological Measurement*, 40(6), 2019.

[15] Autonomio Talos Computer software.

[16] P. Thamarai and K. Adalarasu. An effective method to denoiseeeg, ECG and PPG signals based on meyer wavelet transform. *International Journal of Engineering and Technology(UAE)*, 7(3.34 Special Issue 34):678–683, 2018.

[17] Geoffrey H Tison, José M Sanchez, Brandon Ballinger, Avesh Singh, Jeffrey E Olgin, Mark J Pletcher, Eric Vittinghoff, Emily S Lee, Shannon M Fan, Rachel A Gladstone, Carlos Mikell, Nimit Sohoni, Johnson Hsieh, and Gregory M Marcus. Passive Detection of Atrial Fibrillation Using a Commercially Available Smartwatch. *JAMA cardiology*, 0124:1–8, 2018.

[18] Yifan Zhang, Shuang Song, Rik Vullings, Dwaipayan Biswas, Neide Simões-Capela, Nick Van Helleputte, Chris Van Hoof, and Willemijn Groenendaal. Motion artifact reduction for wrist-worn photoplethysmograph sensors based on different wavelengths. *Sensors (Switzerland)*, 19(3), 2019.