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

RMSSD Estimation From Photoplethysmography and Accelerometer Signals Using a Deep Convolutional Network

Christodoulos Kechris[1] and Anastasios Delopoulos[1]

Abstract— **Heart Rate Variability is a significant indicator of the Autonomic Neural System's functioning, traditionally evaluated from electrocardiogram recordings. Photoplethysmography sensors, like electrocardiograph devices, track the heart's activity and have been widely popularized by their use in smart watches and fitness trackers. In this study we develop a deep learning based approach which is able to successfully estimate the patient's *Root Mean Square of the Successive Differences*, a common heart rate variability metric, from lower quality, less expensive photoplethysmography sensors under a wide range of conditions.**

I. INTRODUCTION

Heart Rate Variability (HRV) is a significant indicator of the Autonomic Neural System's healthy activity, traditionally calculated after analysing electrocardiograph (ECG) recordings. Its non-invasive nature and its ease of calculation have made it popular as it has been used in a plethora of clinical applications like risk assessment after acute myocardial infarction, early detection of diabetic neuropathy [1], prediction of atrial fibrillation [2] and chronic coronary syndrome [3] and others [1]. More recently, with the global spread of COVID-19, research has been conducted on examining whether HRV analysis can be used in order to predict COVID-19 infections and their severity [4]. Further, it also provides useful information about the Autonomic Neural System's behavior during sleep [5] and mental stress levels [6].

Photoplethysmographic (PPG) sensors are optical sensors, worn on the surface of the skin, that can track the blood volume changes caused by the cardiac cycle and thus monitor the heart's activity. Although the sensor's signal is correlated to the ECG, it is considered of lower quality since it can be significantly affected by a plethora of factors. This constitutes the medical analysis of PPG signals difficult.

However PPG sensors demonstrate a lot of advantages. They are a low-cost, easy-to-use alternative to the electrocardiograph. The portable version of electrcardiographs (Holter devices), suffer from significant limitations. They can record data for up to two weeks ([7]) and require a visit to a medical professional before and after the examination. Furthermore, patient's comfortability is an issue since adhesive leads must be attached on the patient's body for the duration of the medical examination.

Popularized by their use in smart-watches and fitness trackers, PPG sensors are available to the general public and can be easily placed on the patient by themselves. Their form factor, usually worn on the wrist as a watch or a band, guarantees a comfortable experience for the patient. Having been widely integrated in smart devices, they cooperate with smartphones and data can be easily retrieved from a smart application. PPG's advantages make it not only a tempting alternative to the Holter and other traditional ECG devices but also open up new possibilities for remote long term continuous medical monitoring, personalized medicine and early diagnosis.

PPG sensors' recordings are of less quality compared to ECG signals. That in combination with the heart rate variability's sensitivity to noise render accurate heart rate variability analysis from PPG signals a challenging task especially in non-sterile, in-the-wild environments where conditions are not optimal. In this study, a Deep Convolutional Neural network is employed in order to accurately estimate the Heart Rate Variability as the *Root Mean Square of the Successive Differences* (RMSSD) from noisy photoplethysmographic signals that are recorded in real-life situations.

II. HEART RATE VARIABILITY METHODS

A. THE HEART RATE VARIABILITY ANALYSIS

Traditionally the HRV is calculated from a segment of ECG recording of appropriate length and the calculation begins with the detection of the segment's R-peaks, since they determine the duration of each cardiac cycle. The timeseries of the time intervals between successive R-peaks, the RR timeseries, is calculated as:

$$RR(i) = t_{R_i} - t_{R_{i-1}} \tag{1}$$

where the i-th R-peak is observed at time instant t_{R_i}. The HRV is calculated after statistical, time or frequency domain analysis on the RR timeseries, which have been standardized by [1]. The duration of the recording segment from which a single HRV value is calculated is also a significant parameter that needs to be determined.

B. HRV ESTIMATION FROM PHOTOPLETHYSMOGRAPHIC SIGNALS

Under optimal conditions HRV can be successfully calculated from PPG signals [8]. However mental or physical stress reduces the agreement between ECG and PPG estimated HRV. This disagreement is attributed to both physiological phenomena, like respiration or arterial stifneess, as well as to external factors like motion.

Sensor motion can introduce motion artifacts (MA) to the PPG signal heavily damaging it and reducing the prominence of the cardiac component. Motion Artifacts are attributed

[1]Multimedia Understanding Group, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece

to sensor movement, tissue deformation and hemodynamic effects as commented by [9]. The motion's frequencies most of the time lie inside the cardiac frequency range, rendering a simple filtering approach obsolete [10].

According to [8], although PPG motion artifact reduction methods exist, they have not been employed on HRV estimation applications. The need for developing robust peak detection methods from noisy PPG signals is also highlighted in [11]. To our knowledge, most motion artifact reduction approaches have been developed with the end goal of estimating the Heart Rate and not the HRV.

Tackling the problem of HRV estimation from PPG requires a method that can either directly estimate HRV or detect the PPG's R-peaks and then calculate the HRV from the estimated RR timeseries. In [12] a generic algorithm, Automatic Multiscale-based Peak Detection (AMPD), is developed in order to detect the peaks in noisy periodic and quasi-periodic signals, a category that the PPG signal falls into. The SpaMa algorithm [13] firstly estimates the Heart Rate and based on these estimations it reconstructs the PPG signal. In the final step it detects the peaks of the *clean* signal and performs HRV analysis on the detected RR timeseries. A deep learning approach is developed in [11] where several deep models are utilized in order to detect the peaks of PPG data.

III. PROPOSED METHOD

A. RMSSD CALCULATION

In this work we focus on estimating the HRV with the RMSSD metric. Given the RR timeseries as described in equation (1), the RMSSD, measured in milliseconds, is calculated as:

$$RMSSD = \left(\frac{1}{N} \sum_{1}^{N} (RR(i) - RR(i-1))^2 \right)^{1/2} \quad (2)$$

The five minute short-term HRV was chosen as the standardized short-term HRV segment length [1].

B. INPUT FORMULATION

The model is fed with the input formed from the PPG and acceleration signals. The acceleration is provided as a motion reference in order to help the model distinguish between the cardiac activity and motion components. According to [14], acceleration might not be the optimal choice for this task. However, at the time this research is being conducted, accelerometers are more popular than dual sensor configurations. Hence we opted for the more widely adopted option.

The PPG and a 3D acceleration signal are passed through a bandpass filter in the frequency range from 0.05 Hz to 4.0 Hz ensuring that components outside the heart's activity frequency range are dropped. The Short-Time Fourier Transform (STFT) of the 4 signals is calculated in 8 second sliding windows with a step of 2 seconds. The use of 8 second windows with a 2 second slide is quite popular in relative literature when processing MA infected PPG signals. The 8-second STFT vector is then normalized and a group of 147 consecutive 8-sec windows compose a 5 minute spectrogram. Finally combining all four spectrograms, one for each channel, a 4-channel two-dimensional sample is generated.

C. PROPOSED CONVOLUTION NEURAL NETWORK

After the input is formulated, it is fed as a 4 channel signal into a Convolutional Neural Network (CNN). The CNN is tasked with reducing the effect of Motion Artifacts and at the same time calculating the RMSSD of the segment. Since the PPG spectrogram portrays the progression of the cardiac cycle frequency across the 5 minute duration of the input segment, the CNN is trained with the target goal of learning the function that maps this frequency progression to the selected HRV method.

In this approach, the HRV is calculated directly from the signal's spectral contents. Although such an approach is quite popular in PPG based Heart Rate estimations, to the best of our knowledge this is the first PPG HRV estimation method that does not rely on the intermediate step of PPG peak detection and RR timeseries formulation.

The model's architecture is described in Table I and is based on the work of [15].

TABLE I
CNN MODEL

Layer	No. Filters	Output Shape	No. Parameters
Convolution	8	(40, 147, 8)	40
Convolution	16	(40, 147, 16)	1168
MaxPooling	-	(20, 74, 16)	-
Convolution	32	(20, 74, 32)	4640
MaxPooling	-	(10, 37, 32)	-
Convolution	64	(10, 37, 64)	18496
MaxPooling	-	(5, 19, 64)	-
Convolution	128	(5, 19, 128)	18496
MaxPooling	-	(3, 10, 128)	-
Convolution	32	(3, 10, 32)	4128
Flatten	-	(960)	-
Dense	256	(256)	246016
Dropout	-	(256)	-
Dense	1	(1)	257

The ELU [16] is used as the activation function in all the layers except from the last one where a linear activation is used. The same layer weight initialization method was followed as proposed by [16].

The robustness of the method's estimations is enhanced by formulating a ten model ensemble. All the models share the same architecture however the initial conditions as well as the optimization path during learning is different for each model. The training set is also randomly shuffled for the training of each of the ensemble's models. The final RMSSD estimations are calculated as the average of all 10 outputs.

IV. EXPERIMENTAL EVALUATION

A. DATASETS

Three publicly available datasets were used to train and evaluate the proposed CNN. All three studies recorded the

PPG signal along with the acceleration of the wrist. The ECG is also provided and it is synchronized with the PPG and the acceleration.

The first one is WESAD [17], a dataset aimed to study the efficacy of inferring a subject's affective state from its physiological signals. Fifteen subjects participated in the study and the experiment for each one lasted approximately two hours. The duration of the experiment was split into 5 stages, during which the neutral mental state, amusement and stress were elicited. The second dataset used is PPGDalia, [15], bearing several similarities to the WESAD dataset. However, here emphasis is given on the subjects performing a wider range of physical activities aiming to simulate real-life situations. PPGDalia also used fifteen participants with approximately a two and a half hour session for each participant. The third dataset is the 2015 IEEE Signal Processing Cup Training dataset [18] consisting of 12 5-minute recordings from 12 subjects which were asked to exercise on a treadmill at speeds from 1km/h to 15km/h.

Employing the three datasets in this study allows for the thorough examination of the effects of both mental and physical stresses on the accuracy of the model's estimations.

B. EXPERIMENTAL SETUP

The model is trained and evaluated on WESAD and PPGDalia while the IEEE dataset is utilized for comparing it to other methods during the evaluation stage.

For the first two datasets, the synchronized ECG, PPG and acceleration are split into 5 minute samples with a 10 second step between two successive samples. This overlay is chosen as a data augmentation strategy in order to generate the appropriate amount of samples needed to train the network. The IEEE data are already split in 12 5-min samples.

From the ECG segment we formulate the RR peak interval timeseries. The PPGDalia dataset provides the locations of the R-peaks, while the Biosppy Library [19] was used for the R-peak detection in the WESAD and IEEE datasets. The ground truth RMSSD of each 5-minute segment is calculated according to the equation (2) while the corresponding input samples are created following the process described in Section III.

The network is trained and evaluated on the PPGDalia and WESAD datasets using the Leave-One-Out (LOO) strategy. This way the network's generalization ability among different, previously unseen patients, can be evaluated. For the formation of the ensemble the training process is repeated 10 times resulting in an ensemble of 10 trained models.

The performance on the IEEE Dataset is evaluated as the average performance achieved by the ensembles that were trained during the LOO training on the PPGDalia.

C. RESULTS

Small errors were achieved on both WESAD and PPGDalia experiments. For WESAD, across all 15 sessions the average MAE was 8.97 ms, the STDE 5.15 ms and the average RMSE was 10.42 ms. Similar error levels were observed for the PPGDalia dataset (7.81 ms MAE, 5.56 ms STDE and 9.66 ms RMSE).

An example of the model's estimations is given in Fig. 1, where WESAD's S10 and PPGDalia's S11 ground truth and estimated RMSSD values are depicted. It is evident that estimated RMSSD closely follows the ground truth values throughout all the session's activities in both cases.

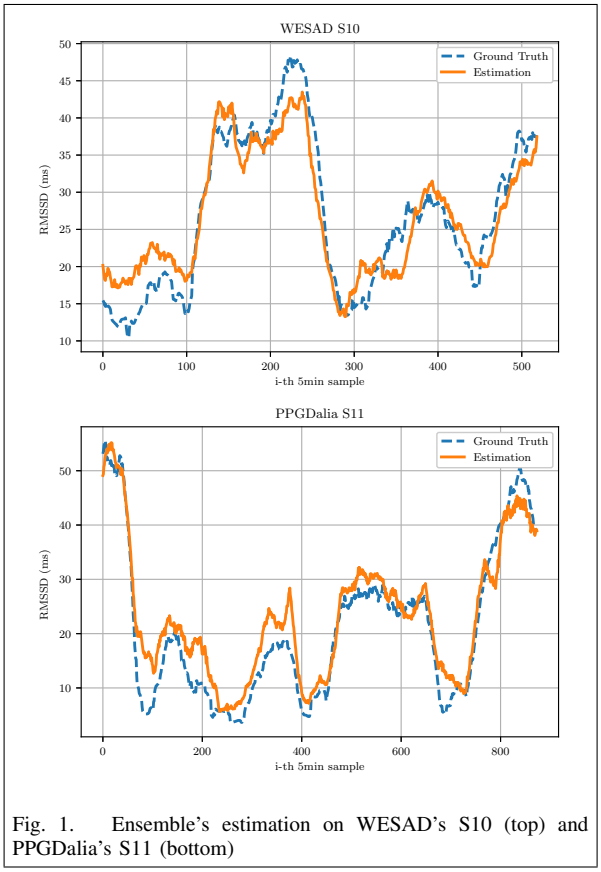

Fig. 1. Ensemble's estimation on WESAD's S10 (top) and PPGDalia's S11 (bottom)

It is interesting examining the estimations' error across the different subjects' mental and physical stress levels (Table II). For each activity, the RMSE, MAE and STDE was averaged across all subjects. The activities' details can be found in [17] and [15] for the two datasets accordingly. We note here that Medi 1 and 2 are meditation stages and the Transient activity is the duration between the ending of an activity and the beginning of the next one. The error remains low during all the different mental stress levels as the RMSE is observed in the range of 8-10 ms, close to the Average RMSE of all the sessions. Small error and error fluctuations are observed throughout the PPGDalia's activities too.

Finally, we compare our approach to [11] and [13], both of which rely on the intermediate step of detecting the PPG's peaks. According to its authors, [11] managed to outperform the AMPD algorithm. Since the IEEE dataset is used, a direct performance comparison is possible. Although the 1-min estimations cannot be exactly compared to 5-min, we report that our approach achieved an order of magnitude more accurate estimations than the method of [11]. In particular the authors report an average error of -63.78 ms with a 65.13

TABLE II

AVERAGED PERFORMANCE ACROSS EACH EXPERIMENT'S STAGE.

WESAD			
Activity	RMSE(ms)	MAE(ms)	STDE(ms)
Base	10.81	9.41	5.11
Stress	8.84	7.87	3.75
Medi 1	10.14	9.39	3.6
Amusement	9.31	8.74	2.95
Medi 2	9.19	8.79	2.5
PPGDalia			
Activity	RMSE(ms)	MAE(ms)	STDE(ms)
Transient	7.11	6.00	3.64
Sitting	11.19	10.31	3.69
Stairs	8.77	8.09	3.19
Table soccer	9.08	8.55	2.92
Cycling	6.15	5.60	2.29
Driving car	7.42	6.88	2.49
Lunch break	9.07	8.12	3.78
Walking	10.66	10.09	3.20
Working	10.65	9.80	3.96

ms standard deviation. In comparison our approach achieved an average error and standard deviation of -6.99 ms and 2.21 ms accordingly.

SpaMa estimates the RMSSD on the 9-min recordings of the Chon Lab dataset ([13]), which bears many similarities to the IEEE dataset in its experimental setup. Although a direct performance comparison between different datasets is not straightforward, or even possible in some cases, the similarities between the two experiments seem to allow a qualitative comparison. Furthermore, the Chon Lab dataset utilizes a forehead mounted sensor, which generally results in PPG recordings of greater quality and less motion artifact interference [20]. Both datasets used similar PPG sensor wave lengths. Across all 10 subjects SpaMa achieved a MAE of 12.23 ms and STDE 9.33 ms. Our approach achieved in the IEEE dataset a MAE of 6.99 ms and STDE of 2.21 ms.

V. CONCLUSIONS

In this work a deep learning approach was developed in order to tackle the problem of accurate RMSSD estimation in real-life conditions from Motion-Artifact affected PPG signals and 3D acceleration as a motion reference. In our experiments the model's accuracy was tested under mental (WESAD dataset) and physical (PPGDalia dataset) stress, conditions which can have a great impact on the agreement between the ECG and the PPG signals and damage the accuracy of PPG estimated HRV. Across the PPGDalia's simulated daily routine the model achieved satisfactory error levels, while it also successfully performed throughout WESAD's mental stress levels. Finally, the proposed approach seems to outperform current state of the art, demonstrating at the same time potential superiority of the direct RMSSD estimation approach.

REFERENCES

[1] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. Cohen, P. Coumel, E. Fallen, H. Kennedy, R. Kleiger *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," 1996.

[2] J. D. Sluyter, C. A. Camargo Jr, A. Lowe, and R. K. Scragg, "Pulse rate variability predicts atrial fibrillation and cerebrovascular events in a large, population-based cohort," *International journal of cardiology*, vol. 275, pp. 83–88, 2019.

[3] J. T. Bigger Jr, J. L. Fleiss, R. C. Steinman, L. M. Rolnitzky, W. J. Schneider, and P. K. Stein, "Rr variability in healthy, middle-aged persons compared with patients with chronic coronary heart disease or recent acute myocardial infarction," *Circulation*, vol. 91, no. 7, pp. 1936–1943, 1995.

[4] A. Natarajan, H.-W. Su, and C. Heneghan, "Assessment of physiological signs associated with covid-19 measured using wearable devices," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–8, 2020.

[5] E. Tobaldini, L. Nobili, S. Strada, K. R. Casali, A. Braghiroli, and N. Montano, "Heart rate variability in normal and pathological sleep," *Frontiers in physiology*, vol. 4, p. 294, 2013.

[6] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel, "Influence of mental stress on heart rate and heart rate variability," in *4th European conference of the international federation for medical and biological engineering*. Springer, 2009, pp. 1366–1369.

[7] A. Galli, F. Ambrosini, and F. Lombardi, "Holter monitoring and loop recorders: from research to clinical practice," *Arrhythmia & electrophysiology review*, vol. 5, no. 2, p. 136, 2016.

[8] A. Schäfer and J. Vagedes, "How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram," *International journal of cardiology*, vol. 166, no. 1, pp. 15–29, 2013.

[9] A.-M. Tăuţan, A. Young, E. Wentink, and F. Wieringa, "Characterization and reduction of motion artifacts in photoplethysmographic signals from a wrist-worn device," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 6146–6149.

[10] F. Wadehn, Y. Zhao, and H.-A. Loeliger, "Heart rate estimation in photoplethysmogram signals using nonlinear model-based preprocessing," in *2015 Computing in Cardiology Conference (CinC)*. IEEE, 2015, pp. 633–636.

[11] T. Wittenberg, R. Koch, N. Pfeiffer, N. Lang, M. Struck, O. Amft, and B. Eskofier, "Evaluation of hrv estimation algorithms from ppg data using neural networks," *Current Directions in Biomedical Engineering*, vol. 6, no. 3, pp. 505–509, 2020.

[12] F. Scholkmann, J. Boss, and M. Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals," *Algorithms*, vol. 5, no. 4, pp. 588–603, 2012.

[13] S. Salehizadeh, D. Dao, J. Bolkhovsky, C. Cho, Y. Mendelson, and K. H. Chon, "A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intense physical activities using a wearable photoplethysmogram sensor," *Sensors*, vol. 16, no. 1, p. 10, 2016.

[14] Y. Zhang, S. Song, R. Vullings, D. Biswas, N. Simões-Capela, N. Van Helleputte, C. Van Hoof, and W. Groenendaal, "Motion artifact reduction for wrist-worn photoplethysmograph sensors based on different wavelengths," *Sensors*, vol. 19, no. 3, p. 673, 2019.

[15] 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.

[16] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[17] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.

[18] Z. Zhang, Z. Pi, and B. Liu, "Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Transactions on biomedical engineering*, vol. 62, no. 2, pp. 522–531, 2014.

[19] C. Carreiras, A. P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred *et al.*, "BioSPPy: Biosignal processing in Python," 2015–, [Online; accessed 11/02/2020]. [Online]. Available: https://github.com/PIA-Group/BioSPPy/

[20] Y. Lee, H. Han, and J. Kim, "Influence of motion artifacts on photoplethysmographic signals for measuring pulse rates," in *2008 International Conference on Control, Automation and Systems*. IEEE, 2008, pp. 962–965.