Draft Report: Deep PPG for Better Heart Rate Estimation

D. Longitudinal predictions on ICU data

Andrei Burghelea, David Gutierrez, Denizhan Kara, and Sung Yoo Kim

***Abstract*—Photoplethysmography (PPG) is a low-cost, non-invasive, and optical technique using an infrared light to measure the volumetric variations of blood circulation in microvascular tissue from the skin surface. Improvements to PPG have brought heart rate measurements to wearable devices such as smartwatches and fitness trackers. Inferring cardiac information (e.g. heart rate) from PPG traces in the context of activity levels beyond the sedentary is extremely challenging, because of interferences caused by motion artifacts (MA). The work undertaken in this project is one of Predictive Modeling - we attempt to predict the heart rate using the time-frequency spectra of synchronized PPG and accelerometer signals as input. The goal of this project is to use the novel large-scale dataset PPG-DaLiA introduced in [1], which includes a wide range of activities performed under close to real-life conditions, to attempt to reproduce the model used in the original reference paper, as well as potentially introduce some improvements using novel techniques.**

I. Introduction

Electrocardiography (ECG) is currently the gold standard for heart rate measurement in clinical settings. More recently, improvements to photoplethysmography (PPG) have brought heart rate measurements to wearable devices, most notably in smartwatches such as the Apple Watch and fitness trackers like FitBit and Garmin devices. The technology has relatively accurate heart rate information to become commonly available in both clinical and everyday environments. Wearable devices are a large and growing segment of consumer devices with demand for increasingly advanced health signals fuelling the growth of more advanced sensor technology. PPG has enabled heart rate calculation in both rested and active states for the average consumer at a significantly reduced cost over ECGs.

PPG has been traditionally used in a variety of devices with application domains ranging from medical to fitness. However, inferring cardiac information (e.g. heart rate) from PPG traces in the context of activity levels beyond the sedentary is extremely challenging, because of interferences caused by motion artifacts (MA). MAs are generally caused by the movement of the sensor module relative to the skin and affect the signal quality and the extraction of the parameters of interest (e.g. heart rate). Recently there have been a number of techniques developed for detecting, removing, or attenuating MA and estimating heart rate, including some based on artificial neural networks ([1]–[6])). Reference [7] provides a comprehensive review of the state-of-the-art research on heart rate estimation from wrist-worn PPG signals.

II. Related work

Certain algorithms explored in the reference [1] including SpaMa, SpaMaPlus, and Schaeck2017 are leveraged to try to eliminate noise in the PPG spectrum data. SpaMa can help eliminate sudden changes due to motion, SpaMaPus can help not carry over fluctuations or error into the next tracking periods, and Schaeck2017 can help decrease the noise with the possible use of multiple channels. These are some algorithms explored in the paper for the data processing that we will explore with potential others in our final report. Reference [3] details the importance of human activity recognition, which is currently achievable using accompanying accelerometer data with high degrees of accuracy and precision. Recent developments of techniques for detecting, removing, or attenuating MA and estimating heart rate include the use of hybrid deep neural networks combining convolutional neural network (CNN) layers with long short-term memory layers (LSTM) in frameworks such as CorNET [4] and PP-Net [5]. This joint framework of CNN and LSTM combined with feature extraction provides a more accurate performance for heart rate estimation. Yet another recent approach is the DeepHeart framework [6], which uses a deep CNN ensemble for denoising PPG signals and removal of MA artifacts, combined with online analysis of the PPG spectrum. The performance of the approaches in [4]–[6] suggests that there is potential for improving the performance of the model in [1] by modifying the network architecture and the input processing.

III. Problem formulation

The work undertaken in this project is one of Predictive Modeling - we attempted to predict the heart rate using the time-frequency spectra of synchronized PPG and accelerometer signals as input. Our aim was to replicate the paper titled Deep PPG: Large Scale Heart Rate Estimation with Convolutional Neural Networks [1]. The goal is to use the novel large-scale dataset introduced in [1], which includes a wide range of activities performed under close to real-life conditions, to attempt to reproduce at least one model used in the original reference paper, as well as potentially introduce some improvements using Long short-term memory (LSTM) and other novel techniques. The dataset has a total of 11 attributes with over 8 million instances. This data includes raw sensor data with two devices RespiBan (chest-worn) and Empatica E4 (wrist-worn). The RespiBan provides the ECG, breathing, and motion signals. The Empatica provides additional data including BVP, EDA, and body temperature and motion which will be used as additional features.

IV. Methodology

Reference [1] provides the details about the data collection protocol. We have used the same steps as identified in the reference paper. This means we have segmented the time-series data with a sliding window of length 8 seconds and a window shift of 2 seconds. We consider the first PPG channel and all three accelerometer channels. We may consider additional PPG channels as input for our deep learning model if model development and timing allow. As a second step, we apply Fast Fourier Transform (FFT) on each time-series segment. The result of this step is Nch = 4 channel time-frequency spectra, one per signal channel. In the next step, we cut these spectra, keeping only the 0–4 Hz interval (4 Hz corresponds to 240 bpm). The resulting number of FFT points per segment and channel is NFFT=1025. Finally, z-normalization (zero mean and unit variance) is performed on each channel's spectrum. The final Nch time-frequency spectra serve as input for the deep learning model.

For the data preparation stage, we segmented the data with 8 seconds of sliding windows with a time rolling mechanism with 2 seconds shifts. This enabled us to extract features that we need from the data. Since we need both the time and frequency domain behavior of sensor channels, we extracted both time and frequency domain characteristics of each window. Time-domain characteristics include features indicating regularities and statistical measures on windows like mean, max, standard deviation, and unique/recurring value count, while frequency-domain features represent more complex, yet periodic information within each window. These features are planned as FFT coefficients, number of CWT (continuous wavelet transform) peaks, and autoregression of each window.

Since we were unsure about which features or characteristics will be most contrastive towards heart-rate estimation, we ended up extracting a large number of features from each window that we think might represent information towards modeling. However, this approach did lead to some correlation between features as well as some features that represent no information, which can be considered as noise. For this reason, we also applied univariate and multivariate feature selection techniques, which eliminated high correlation and noise features.

After the selection procedures, we started with a variety of modeling procedures. We made use of various models like deep learning models as well as simpler options like tree-based models, which would allow us to perform benchmarking on our additional features with easier training procedures. Moreover, such models let us understand the information value of the considered features without being much affected by scaling and correlation issues between input features. Here, our goal is to determine which features are beneficial for our task as well as which architecture is likely to give us the best results.

For the initial training, we have extracted both time and frequency domain features from the preprocessed 8 second time windows and benchmarked them on a subject basis. The features include simple extractions like mean, maximum, and standard deviation of acceleration values in each window, which are intended to capture the time domain characteristics of the features, while more complex features like FFT coefficients, autocorrelation, and linear trend capture frequency domain characteristics of the signals, which are especially important to predict heart rate in the presence of high-frequency acceleration dynamics, which are very common during movements and sports activities for each subject.

In parallel to the improved features, we intend to implement the original CNN model from [1]. This will let us both implement the paper model itself, which is the minimum requirement for the project task, but also let us build on and improve the model itself with both architectural changes as well as new features, which we have experimented on in the initial training.

The performance metric is the same as the one used in [1], the mean absolute error (MAE) defined as:

$$MAE = \frac{1}{W} \sum_{w=1}^{W} BPM_{est}(w) - BPM_{ref}(w) \quad (1)$$

V. Experiments

Our target is to achieve a lower or equal MAE than the reference paper [1] on 6/9 Activities.

Our initial training results on a subject-wise training model yielded an MAE between 1.6 and 2.8, which is a significant improvement over the MAE recorded in the reference paper [1]. This improvement was the result of a simple gradient boosting model (Xgboost) without hyperparameter tuning. The evaluation was done by splitting each subject's data on a standard 80:10:10 train-test-validate basis with 80% of the data used for training. The initial results are presented in Table I.

On our second part we will build a convolutional neural network (CNN) to incorporate deep learning on activity-based training. Here, the goal is to replicate the CNN model proposed in Reference [1]. The reference model has a deep CNN architecture with two initial convolution layers and up to 8 additional convolution-maxpool layers. The first two layers are designed to fuse first the input channels (accelerometer measurements and PPG), and the segments used in heart rate tracking, respectively. The successive convolution-maxpool layers are designed to increase the learning effectiveness of the model. One last convolutional layer is included in the model in order to reduce the input dimension of the fully connected layer. The last two layers are fully connected, with the first flattening the input and the second one providing a single value estimated heart rate. The model used an exponential linear unit (ELU) as activation function for all the layers. The loss function is defined as the absolute difference between the estimated value for a given segment and the corresponding ground truth. This model is a regression problem. For optimization the model uses the Adam optimizer [8]. For the model we are trying to replicate, we are going to be including the same filter size to decrease the spatial size of the output. This will decrease the number of features volume but increase the depth of the volume. The paper mentioned that this technique has been successfully implemented in other architectures like VGGNet, ResNet, and DenseNet [1].

At the time of this preliminary report this model is still under development. We currently have the model created and the training implemented. However, the evaluation of the model to a validation dataset with error with MAE needs to be done and the data loading with our datasets. Following our timeline, we will have implemented the initial network soon and be starting on updating our model based on features manipulation or network design.

Our next steps include the use of additional features we benchmarked in the initial training, which have performed very well. Also, we are planning to go with an ensemble architecture that includes a simple activity recognition model. The activity is very important in the HR estimation and it is usually not explicitly available as an input to the model, thus we are planning to incorporate it for a future improvement, which is proposed in some other works as well [3]. Moreover, we are aware that the interpretability aspect of the model is crucial for both improving and approval of it. We have performed a literature search on the interpretability aspect and found the SHAP algorithm [9], which

TABLE I
INITIAL RESULTS

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	All
Reference CNN Average	8.45	7.92	5.96	7.86	18.97	13.55	5.16	11.49	10.65	6.07	9.87	9.95	5.25	5.85	5.25	8.82 ± 3.8
Reference CNN Ensemble	7.73	6.74	4.03	5.9	18.51	12.88	3.91	10.87	8.79	4.03	9.22	9.35	4.29	4.37	4.17	7.65 ± 4.2
Gradient Boosting (Xgboost)	2.32	2.1	2.3	2.22	2.13	2.43	2.58	2.29	2.11	1.94	2.85	1.64	2.7	2.83	2.03	2.30 ± 0.34

can help us with the interpretation aspect of the model.

VI. CONCLUSION

In this paper, we sought to improve over state-of-the-art heart rate prediction algorithms. One of the primary innovations to this end was the preprocessing of various non-linear statistical features of the time-segmented windows of signals in order to get an equal or lower MAE score during evaluation of our model. The preprocessed features took time to implement due to the size of the raw data and the filtering of the discrepancies that occur. The preprocessed data included FFT coefficients, autocorrelation, and linear trend capture frequency domain characteristics. Using the preprocessed features in a given window size, we were able to prove that the features were beneficial using a preliminary testing with XGBoost. The next steps include finishing up the recreated CNN model [1], adjusting our input features, and trying other deep learning architectures to lower the MAE score. Currently, we have done a lot of work in our feature selection and data processing that will allow for us to easily input them into our models. We will soon be implementing the initial model and adjusting it for better results.

REFERENCES

[1] 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.

[2] 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.

[3] E. Brophy, W. Muehlhausen, A. F. Smeaton, and T. E. Ward, "Optimised convolutional neural networks for heart rate estimation and human activity recognition in wrist worn sensing applications," *arXiv preprint arXiv:2004.00505*, 2020.

[4] D. Biswas, L. Everson, M. Liu, M. Panwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Acharyya, C. Van Hoof, M. Konijnenburg *et al.*, "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.

[5] M. Panwar, A. Gautam, D. Biswas, and A. Acharyya, "Pp-net: A deep learning framework for ppg-based blood pressure and heart rate estimation," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 000–10 011, 2020.

[6] X. Chang, G. Li, G. Xing, K. Zhu, and L. Tu, "Deepheart: A deep learning approach for accurate heart rate estimation from ppg signals," *ACM Transactions on Sensor Networks (TOSN)*, vol. 17, no. 2, pp. 1–18, 2021.

[7] D. Biswas, N. Simões-Capela, C. Van Hoof, and N. Van Helleputte, "Heart rate estimation from wrist-worn photoplethysmography: A review," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6560–6570, 2019.

[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[9] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.