IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 7, 2020, accepted March 16, 2020, date of publication March 19, 2020, date of current version April 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2981956

Deep Learning for Heart Rate Estimation From Reflectance Photoplethysmography With Acceleration Power Spectrum and Acceleration Intensity

HEEWON CHUNG, HOON KO, HOOSEOK LEE, AND JINSEOK LEE, (Senior Member, IEEE)
Department of Biomedical Engineering, Wonkwang University College of Medicine, Iksan 54538, South Korea

Corresponding author: Jinseok Lee (gonasago@wku.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF), Ministry of Science, ICT, and Future Planning, through the Basic Science Research Program under Grant NRF-2020R1A2C1014829 and Grant NRF-2015M3A9D7067215.

ABSTRACT *Objective:* A wearable reflectance-type photoplethysmography (PPG) sensor can be incorporated in a watch or band to provide instantaneous heart rates (HRs) with minimum inconvenience to users. However, the sensor is sensitive to motion artifacts (MAs), which results in inaccurate HR estimation. To address this problem, we propose a new neural network for deep learning to ensure accurate HR estimation even during intensive exercise. *Methods*: We propose a new deep neural network based on multiclass and non-uniform multilabel classification for HR estimation. It comprises of two convolutional layers, two long short-term memory (LSTM) layers, one concatenation layer, and three fully connected layers including a softmax. The proposed model feeds the power spectra from the PPG and acceleration signals along with the acceleration intensity to the input layer. We also present a new scheme to evaluate the loss value by modifying the true HR value into a Gaussian distribution. *Results*: We used 48 training datasets and evaluated 23 isolated testing datasets. The proposed model exhibited average absolute error of less than 1.5 bpm for all the training and test datasets—1.09 bpm for the training dataset and 1.46 bpm for the test dataset. *Conclusion*: The proposed model outperforms the state-of-the-art methods for accurate estimation of HR. *Significance*: It precisely estimates the HRs with robustness even during intensive physical exercise, as evidenced by the accuracy when PPG signals are severely corrupted by MAs.

INDEX TERMS Reflectance-type photoplethysmography, instantaneous heart rate, deep learning, convolutional layer, long short-term memory (LSTM) layer.

I. INTRODUCTION

Reflectance-type photoplethysmography (PPG) sensor measures intensity changes in the light reflected from skin, providing PPG signals that represent the changes in the arterial blood volume between the systolic and diastolic phases of a cardiac cycle. The sensor has gained attention because it can be incorporated in watches or bands to measure and monitor instantaneous heart rates (HRs), which minimizes inconvenience to users. However, the sensor is sensitive to motion artifacts (MAs), which originate from pressure and movement applied on the wrist on which the PPG sensor is worn. The MAs eventually result in inaccurate HR estimation. A few years ago, Zhang *et al.* shared the datasets containing simultaneously measured acceleration and PPG signals during exercise [1], which has prompted research on MA cancelation in PPG sensors using acceleration signals. Current state-of-the-art methods have reached the accuracy of 2–3 beats per minute (bpm) on average absolute errors (AAEs) during intensive exercise [1]–[13]. Most state-of-the-art methods estimated the HR using the two main stages of MA cancelation and HR tracking. For MA cancelation, they considered the power spectrum from the simultaneously measured acceleration signal as motion artifacts (MAs), and removed or attenuated the power from the PPG power spectrum. For HR tracking, they exploited the assumption that the change in the HR between two consecutive segments is not significant, and predicted or corrected the HR results. Various

The associate editor coordinating the review of this manuscript and approving it for publication was Alessio Vecchio.

signal processing algorithms such as high-resolution spectra including variable frequency complex demodulation [1]–[3], adaptive filters including a Wiener filter [4]–[8], decomposition including ensemble empirical mode decomposition, singular value decomposition [9], [10], and non-linear filters including Kalman filters (KF) and particle filters [11]–[13] have been adopted for MA cancelation or/and HR tracking. Nevertheless, despite the efforts and advances in the algorithms, the methods did not always provide accurate results.

One of the most challenging issues is low signal-to-noise ratio (SNR) in the PPG signal. Regardless of the method that employs high-resolution spectra for MA cancelation, it is almost impossible to find the accurate HR when the power corresponding to the HR in the measured PPG is very low. This low SNR mainly occurs when a subject is performing intensive physical exercise. The HR tracking approach may minimize the outliers in the results, but it becomes ineffective if low SNR occurs continuously for a long time. To overcome the issue, our group recently presented the finite state machine (FSM) framework, which ignored low-quality signal segments or inaccurate estimation results [14], [15]. We could provide very accurate HR results based on the FSM framework. However, the framework discarded nearly half of the results collected during intensive physical exercise. Later, we presented multi-mode particle filtering (MPF) methods, which could lower the discard rate of the results while preserving the accuracy [16]. Nevertheless, the discard rate of the results during intensive physical exercise was higher than 30%. Very recently, a deep learning approach was considered. In [17], a deep learning framework named CorNet was proposed by modeling convolutional and long short-term memory (LSTM) layers followed by a fully connected layer. In the model, band-pass filtered PPG data were used for an input layer.

In this paper, we propose a new deep neural network based on multiclass and non-uniform multilabel classification for HR estimation. In our proposed model, we consider two power spectra from PPG and acceleration signals for an input layer. In addition, we use the acceleration signal intensity in the input layer. We hypothesize that the acceleration signal intensity can provide information on the change in the HR in the near future: high intensity represents intensive movements, which may change the HR. The proposed model comprises two convolutional layers, two LSTM layers, one concatenation layer, and three fully connected layers including a softmax. In the proposed model, the power spectra from the PPG and acceleration signals are fed into two convolutional layers, which provide MA cancelation in the PPG power spectrum. The outputs are flattened and connected to one fully connected layer, which is subsequently concatenated with the acceleration signal intensity. Then, the outputs are fed to the two LSTM layers followed by the additional fully connected layers that include a softmax. The LSTM layers track the HR tracking with minimum outliers. In the proposed model, we also present a new scheme to evaluate the loss value by modifying the true HR value into a Gaussian distribution. The performance of the proposed model is evaluated by comparison with previously reported results [1]–[17], [22].

II. METHODS

A. DATASETS

We used two datasets to evaluate our proposed model for HR estimation—the IEEE Signal Processing Cup (ISPC) 2015 dataset ($n = 23$) and direct measurements obtained by our developed device ($n = 48$). Both sets of data include multichannel PPG signals and three-axis accelerometer signals simultaneously measured by devices worn on the wrist during intensive physical exercise. For true HR reference, ECG signals were simultaneously measured on the chest.

More specifically, the ISPC dataset includes 5-min two-channel PPG signals and three-axis acceleration signals sampled at 125 Hz for 23 subjects, which are publicly downloadable [18]. The dataset is grouped into three groups: Type 1 (T1), Type 2 (T2), and Type 3 (T3). In the T1 group ($n = 12$), the subjects run on a treadmill at various speeds: 30 s of rest, 1 min at 6–8 km/h, 1 min at 12–15 km/h, 1 min at 6–8 km/h, 1 min at 12–15 km/h, followed by 30 s of rest. In the T2 group ($n = 5$), the subjects perform various actions such as running, jumping, push-ups, shaking hands, stretching, and pushing. In the T3 group ($n = 6$), the subjects perform intensive arm movements such as boxing.

Our data comprise 12-min three-channel PPG signals and three-axis acceleration signals sampled at 50 Hz. The dataset is classified into two groups named BAMI-I and BAMI-II. In the BAMI-I dataset ($n = 25$), the exercise protocol included 1 min of rest, 2 min of walking for warm-up, 3 min of running at 6–8 km/h, 2 min of walking, 3 min of running at 8–12 km/h, and 1 min of walking to cool down. The subjects comprised 10 males and 14 females with average age of 26.9±4.8 years. The entire exercise process was performed on a treadmill. In the BAMI-II dataset ($n = 23$), the exercise protocol included 1 min of rest, 2 min of walking for warm-up at 3–4 km/h, 4 min of running at 6–8 km/h, 4 min of walking at 3–4 km/h, and 1 min of rest to cool down. During every 4-min session of running and walking, the subjects walked or ran while holding a treadmill bar during the last two minutes of the session. We designed the session to reflect cardiac rehabilitation exercise for cardiac patients with poor exercise ability—they normally walk or run by holding a treadmill bar. The subjects comprised 17 males and 6 females with average age of 22.0±1.7 years. The entire exercise process was also performed on a treadmill. For both datasets, the reference true HRs were measured by ECG data simultaneously recorded by a 24-h Holter monitor (SEER Light, GE Healthcare, Milwaukee, WI, USA). All the data were collected at Wonkwang University by trained personnel from June to July 2018 for BAMI-I, and from March to April 2019 for BAMI-II. This study was approved by the Institutional Review Board of Wonkwang University, Republic of Korea (WKUIRB 201805-032-01). All participants

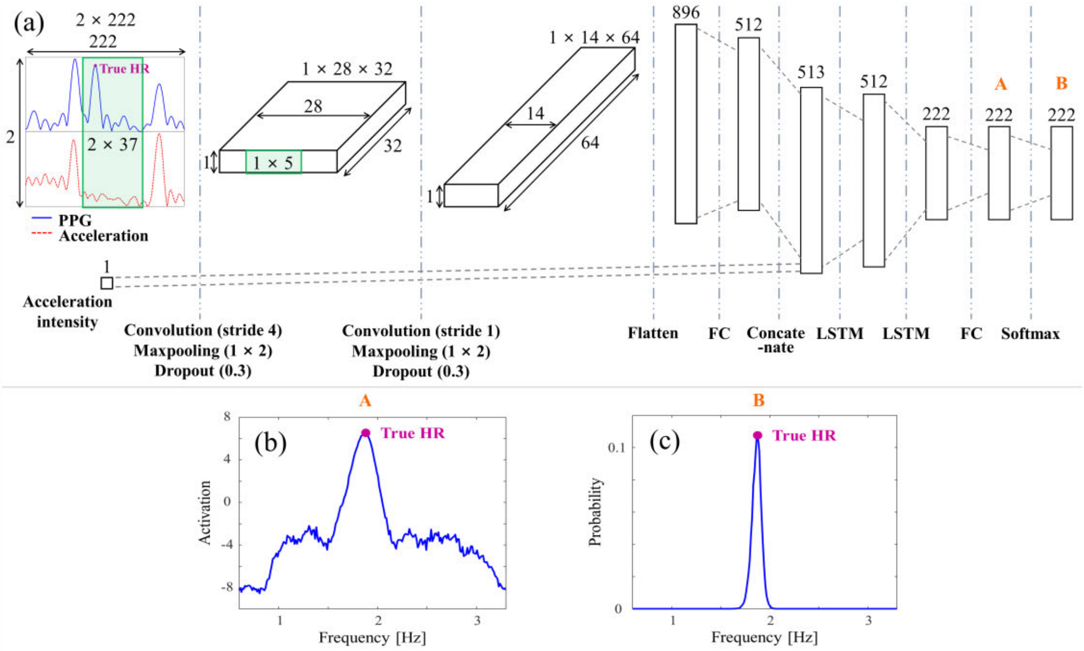


FIGURE 1. Architecture of our proposed model; (a) 2D convolutional layer, 1D convolutional layer, one fully connected layer, two LSTM layers, and one fully connected layer are sequentially structured, (b) 222 activations before a softmax layer, and (c) final output providing the final probabilities for the true HR.

provided their written informed consent. All raw signals in the BAMI-I and -II datasets are publicly downloadable [19].

In this study, we chose the ISPC dataset ($n = 23$) as training data since it includes a variety of actions such as waking, running, resting, jumping, push-ups, shaking hands, stretching, pushing and boxing. We also included our own dataset named BAMI-I ($n = 25$) in the training data to increase the training data size, and tested the trained model using another dataset named BAMI-II ($n = 23$). For all datasets, the ECG-based HRs were calculated using 8-s windows with 2-s shifts (6-s overlap), yielding the HRs for every 2 s. The same window length (8 s) and shift (2 s) were used throughout this study to assess the performance of our proposed model relative to the existing algorithms [1]–[17].

B. PREPROCESSING FOR INPUT LAYER

We denote the estimated and true HRs in the i^{th} window by $HR_{est}(i)$ and $HR_{true}(i)$, respectively, where $i = 1, 2, \ldots I$; I represents the number of 8-s windows. We also denote the measured multichannel PPG signals and three-axis acceleration signals in the i^{th} 8-s window by $\boldsymbol{S}_n(i)$ and $\boldsymbol{A}_m(i)$, respectively, where $n = \{1, 2, \ldots N\}$, N being the number of photosensors, and $m = \{1,2,3\}$ represents the x-, y- and z-axis, respectively. Note that $N = 2$ for the ISPC dataset, and $N = 3$ for BAMI-I and II datasets. We filtered all measured signals of $\boldsymbol{S}_n(i)$ and $\boldsymbol{A}_m(i)$ using a fourth-order Butterworth band pass filter (BPF) with cutoff frequencies of 0.4 and 4 Hz. We then normalized $\boldsymbol{S}_n(i)$ to a zero mean with unit variance and averaged them for the single signal denoted by $\boldsymbol{S}(i)$. We downsampled $\boldsymbol{S}(i)$ and $\boldsymbol{A}_m(i)$ to 25 Hz, which could reduce computational load with little accuracy degradation [9]. After the down-sampling, we computed the power spectra via a 2,048-point Fast Fourier Transform (FFT), where 200 sample vector (8-s data) was padded with trailing zeros to length 2,048, and the frequency bin resolution became 0.012 Hz (0.73 bpm). Subsequently, we normalized the power spectrum with the minimum value of zero and the maximum value of one: $\boldsymbol{P}^S(i)$ from $\boldsymbol{S}(i)$, and $\boldsymbol{P}^A_m(i)$ from $\boldsymbol{A}_m(i)$. We further averaged the power spectra from three-axis acceleration signals: $\boldsymbol{P}^A(i) = \frac{1}{3}\sum_{m=1}^{3}\boldsymbol{P}^A_m(i)$. Given $\boldsymbol{P}^S(i)$ and $\boldsymbol{P}^A(i)$, we extracted only the possible HR range between 0.6 and 3.3 Hz, which resulted in 222 frequency-bin power spectra. Then, each extracted instance of data was of size 1 × 222 with frequency resolution of 0.0122 Hz, which was equivalent to 0.73 bpm. We denote the resultant 1 × 222 size data by $\boldsymbol{P}^s(i)$ and $\boldsymbol{P}^a(i)$, respectively. In addition to the power spectra, we also computed the average of the envelop amplitudes in each 8-s window denoted by $I^a(i)$, which indicates the acceleration intensity. In our proposed model, $\boldsymbol{P}^s(i)$, $\boldsymbol{P}^a(i)$, and $I^a(i)$ are in the input layer.

C. MODEL DESCRIPTION

The architecture of our proposed model is summarized in Fig. 1(a). It contains eight layers: two convolutional, two LSTM, one concatenation, and three fully connected layers including a softmax. More specifically, a 2D convolutional layer, 1D convolutional layer, a flatten layer, a concatenation layer, one fully connected layer, two LSTM layers, and one fully connected layer followed by a softmax are sequentially structured. For the input layer, the power spectra from the

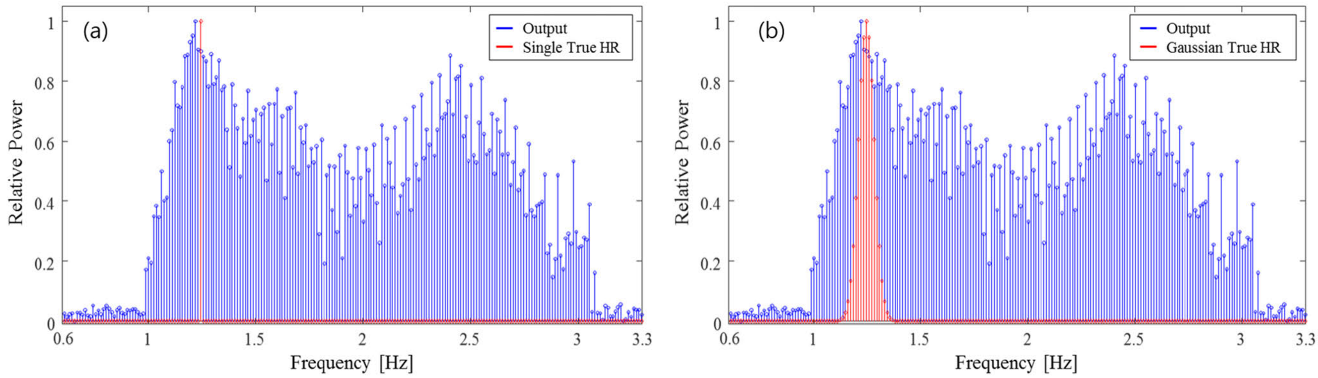


FIGURE 2. True HR probabilities y_c^i with activations before the softmax layer, (b) modified true HR probabilities $\hat{y}_c^i$ with activations before the softmax layer.

PPG and acceleration signals are aligned with the length of two (size 2 × 222): the top signal is from PPG ($\boldsymbol{P}^s(i)$), and the bottom is from the acceleration ($\boldsymbol{P}^a(i)$). Note that the power spectra are based on each 8-s window, which is subsequently shifted by 2-s (6-s overlap). The input layer is fed into the 2D convolutional layer with 32 2 × 37 kernels and stride of 4, followed by the nonlinear function of leaky rectified linear unit (ReLU) and 1 × 2 max pooling with stride of 2. The resultant feature maps with size of 1 × 28×32 represent the intermediate PPG power spectrum with MA cancelation. The feature maps are fed into the 1D convolutional layer with 64 1 × 5 kernels and stride of 1, followed by the leaky ReLU and 1 × 2 max pooling with stride of 2, which provides other feature maps with size of 1 × 14×64 representing the PPG power spectrum with MA cancelation. The resultant feature maps are flattened to 896 nodes, which are fully connected to 512 nodes with leaky ReLU.

Subsequently, the acceleration intensity $I^a(i)$ is concatenated with the 512 activations. We hypothesized that the acceleration intensity indicates how the current HR value will change in the near future. For instance, high acceleration intensity indicates high-intensity motion, which may increase the HR. The 513 concatenated nodes are fed into two LSTM layers. The two LSTM layers act as an HR tracking algorithm by considering the local HR trace pattern. Because of the LSTM layers, the dominant frequency corresponding to the HR is not severely deviated in consecutive windows even when the signal-to-noise ratio (SNR) of the PPG signal is extremely low. The first LSTM layer includes 512 nodes with 6 timesteps, all of which are connected to the second LSTM layer. Note that we found that 6 timesteps provided the best accuracy. The numerical analysis of the effect of the timestep length is presented in Section III. The second LSTM layer also includes 6 timesteps, each of which provides an output with length of 222. Then, only the output from the last timestep is fed into the fully connected layer connecting to 222 nodes with leaky ReLU. Fig. 1(b) shows the resultant 222 activations representing the final PPG power spectrum with MA cancelation in the frequency range of 0.6–3.3 Hz with frequency resolution of 0.012 Hz. The activations are fed into a softmax layer, which provides the final probabilities for the true HR value, as shown in Fig. 1(c).

To avoid overfitting, we applied dropout to the two convolutional layers and two LSTM layers. For the convolutional layers, the dropout rates were set to 0.3. For the LSTM layers, the dropout rates for the linear transformation of inputs were 0.3, and the dropout rates for the recurrent state were 0.2.

D. MULTICLASS AND NON-UNIFORM MULTILABEL CROSS-ENTROPY BASED COST FUNCTION

Given a set of parameters of the model, the softmax provides the probabilities corresponding to each HR value subdivided in the 222 frequency bins: from 0.6 to 3.3 Hz with an interval of 1/222. Thus, with the true HR value, we may consider the multiclass classification problem, which calculates the cost ξ via multiclass cross-entropy as

$$\xi = -\sum_{c=1}^{222} y_c^i log\left(\tilde{y}_c^i\right), \tag{1}$$

where y_c^i and $\tilde{y}_c^i$ are the true HR probability and the predicted HR probability, respectively, for the c^{th} frequency bin at the i^{th} window. In the multiclass cross-entropy, y_c^i has the value of one only when the c^{th} frequency bin corresponds to a true HR. Otherwise, y_c^i has the value of 0. Then, the formulation can be simplified as

$$\xi = -log(\tilde{y}_{c=true}^i), \tag{2}$$

where $\tilde{y}_{c=true}^i$ is the predicted HR probability in the frequency bin covering the true HR value.

However, this approach has a drawback that the frequency bin covering the true HR value may not exactly represent the true HR, as shown in Fig. 2(a), where the true HR value (red) and the dominant activation (blue) are not overlapped. The issue can be attributed to the fact that the ECG- and PPG-based HRs are not exactly overlapped due to the peak morphology of each signal [20], [21]. In addition, different sampling rates also result in slight difference in the determination of the HR value. Note that the true HR was obtained from ECG data sampled at 50 Hz from BAMI I and II datasets and that sampled at 125 Hz from the ISPC dataset whereas the frequency bins were derived from the PPG signal downsampled at 25 Hz. To resolve the issue, we multiplied the cost ξ by the normalized Gaussian function, where a Gaussian distribution with the center as the true HR was normalized to have a maximum value. The modified cost function is

$$\xi = -\log\left(\tilde{y}^{i}_{c=true}\cdot\frac{\exp\left(-\frac{(HR_{true}(i))^2}{2\sigma^2}\right)}{\max\left[\exp\left(-\frac{(HR_{true}(i))^2}{2\sigma^2}\right)\right]}\right) \quad (3)$$

where HR_{true} represents the true HR. From the modification, we could alleviate the non-overlapping issue between the ECG- and PPG-based HRs. Fig. 2(b) shows the modified true HR probabilities inside the parenthesis in (3), which are plotted in red. In this study, we chose the standard deviation $\sigma = 3$. The numerical analysis of the effect of σ is described in Section III.

E. IMPLEMENTATION AND PERFORMANCE EVALUATION

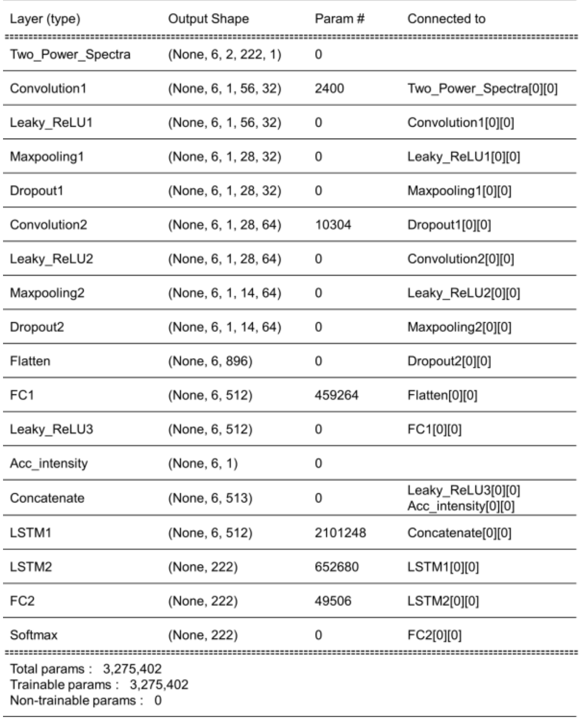
Our proposed model was implemented using Tensorflow package, which provides a Python API for tensor manipulation for deep learning. We also used Keras, which is now the official frontend of Tensorflow. Keras and Tensorflow, in combination with standard Python libraries such as Numpy and Matplotlib, were used to build the model and analyze the results. We trained our model with the ADAM optimizer with a learning rate of 0.0001 and batch size of 1 on NVIDIA GeForce GTX 1080 Ti GPU. Fig. 3 shows the printed textual summary of our proposed model run on Keras. The total number of parameters (weights and biases) was 3,275,402.

For the performance evaluation, four-fold cross-validation was performed in this study to confirm the generalization ability of the proposed model for HR estimation. The training dataset ($n = 48$) was randomly shuffled and divided into four equal groups, each of which included the data for 12 subjects. Subsequently, three groups were selected for training the model and the remaining one group was used for validation. This process was repeated four times by shifting the validation group. Then, we averaged the mean validation costs of the four validation groups according to each epoch and found the optimal epoch that provides the lowest validation cost. Then, we re-trained the model using the entire training dataset ($n = 48$) with the optimal epoch. The isolated test dataset ($n = 23$) was only evaluated after the model was completely trained using the training dataset. This hold-out method provides an unbiased evaluation of the final model fit on the training dataset. To examine the HR estimation, we used the average of absolute errors [AAE (bpm)] and the average of relative absolute errors [ARE (%)].

III. RESULTS

A. RESULTS FROM TRAINING AND TEST SETS

Based on our proposed model, we found that the resultant AAE and ARE values were 1.09 bpm and 0.92% for the training dataset ($n = 48$), and 1.46 bpm and 1.23% for the test dataset ($n = 23$), respectively. The performance is summarized in Table 1. More specifically, the AAE and ARE values were 0.76 bpm and 0.66% with the ISPC dataset, and 1.39 bpm and 1.17% with the BAMI-I dataset, respectively.



```
Layer (type)          Output Shape           Param #    Connected to
==========================================================================
Two_Power_Spectra     (None, 6, 2, 222, 1)   0
Convolution1          (None, 6, 1, 56, 32)   2400       Two_Power_Spectra[0][0]
Leaky_ReLU1           (None, 6, 1, 56, 32)   0          Convolution1[0][0]
Maxpooling1           (None, 6, 1, 28, 32)   0          Leaky_ReLU1[0][0]
Dropout1              (None, 6, 1, 28, 32)   0          Maxpooling1[0][0]
Convolution2          (None, 6, 1, 28, 64)   10304      Dropout1[0][0]
Leaky_ReLU2           (None, 6, 1, 28, 64)   0          Convolution2[0][0]
Maxpooling2           (None, 6, 1, 14, 64)   0          Leaky_ReLU2[0][0]
Dropout2              (None, 6, 1, 14, 64)   0          Maxpooling2[0][0]
Flatten               (None, 6, 896)         0          Dropout2[0][0]
FC1                   (None, 6, 512)         459264     Flatten[0][0]
Leaky_ReLU3           (None, 6, 512)         0          FC1[0][0]
Acc_intensity         (None, 6, 1)           0
Concatenate           (None, 6, 513)         0          Leaky_ReLU3[0][0]
                                                        Acc_intensity[0][0]
LSTM1                 (None, 6, 512)         2101248    Concatenate[0][0]
LSTM2                 (None, 222)            652680     LSTM1[0][0]
FC2                   (None, 222)            49506      LSTM2[0][0]
Softmax               (None, 222)            0          FC2[0][0]
==========================================================================
Total params :  3,275,402
Trainable params :  3,275,402
Non-trainable params :  0
```

FIGURE 3. Printed textual summary of our implemented model run on Keras.

TABLE 1. Performance summary with each dataset: ISPC and BAMI-I were used as training set, and BAMI-II was used as test set.

		AAE (bpm)		ARE (%)	
Training set	ISPC (*n*=23)	0.76	1.09	0.66	0.92
	BAMI-I (*n*=25)	1.39		1.17	
Test set	BAMI-II (*n*=23)	1.46		1.23	

Table 2 compares our results with the 12 previously reported results obtained with the ISPC dataset. Note that some studies reported the results based on only the first 12 recordings; others reported all results except for the 13thsubject, and only a few reported the performance for the entire data pertaining to the 23 subjects. Our proposed model exhibited AAE and ARE of 0.67 bpm and 0.50%, respectively, for the first 12 subjects (*subjects 1–12*; T1 activity). These error values are the lowest in comparison with the other results, where the AAEs and AREs ranged from 1.02–2.34 bpm and 0.81–1.82%, respectively [1-4, 6-8, 11-13, 17, 22]. With the exception of the 13th subject, our algorithm outperformed the existing methods for the entire dataset [1-4, 6-8, 13, 17, 22]. The current state-of-the-art methods exhibited AAE of 1.47–2.73 bpm whereas our method exhibited AAE of 0.75 bpm. In addition, for the

TABLE 2. Comparison of the performances of various methods: ISPC dataset evaluated as a trained data.

Subj.	Act.	TROIKA [1] AAE(ARE) *bpm(%)*	JOSS [2] AAE(ARE) *bpm(%)*	SpaMa [3] AAE(ARE) *bpm(%)*	Spectrap [6] AAE(ARE) *bpm(%)*	WFPV [7] AAE(ARE) *bpm(%)*	Cor NET [17] AAE *bpm*	PARH ELIA [12] AAE *bpm*	PF [11] AAE *bpm*	Fallet [4] AAE(ARE) *bpm(%)*	Galli [13] AAE(ARE) *bpm(%)*	Motin [8] AAE(ARE) *bpm(%)*	Mash hadi [22] AAE *bpm*	Proposed Model AAE(ARE) *bpm(%)*
1	T1	2.29(2.18)	1.33(1.19)	1.23(1.14)	1.18(1.04)	1.25(1.15)	6.23	1.82	2.21	1.75(1.59)	2.72(2.11)	1.18(1.09)	1.81	0.64(0.51)
2	T1	2.19(2.37)	1.75(1.66)	1.59(1.30)	2.42(2.33)	1.41(1.30)	1.83	1.29	1.71	1.94(1.99)	3.25(3.02)	1.65(1.52)	1.44	0.71(0.63)
3	T1	2.00(1.50)	1.47(1.27)	0.57(0.45)	0.86(0.66)	0.71(0.59)	0.89	0.80	1.11	1.17(1.02)	1.40(1.11)	0.75(0.63)	0.63	0.59(0.48)
4	T1	2.15(2.00)	1.48(1.41)	0.44(0.31)	1.38(1.31)	0.97(0.88)	0.49	0.99	1.71	1.67(1.51)	1.21(1.04)	0.87(0.76)	1.16	0.62(0.47)
5	T1	2.01(1.22)	0.69(0.51)	0.47(0.31)	0.92(0.74)	0.75(0.57)	0.40	0.65	1.10	0.95(0.75)	0.93(0.70)	0.74(0.56)	0.83	0.67(0.47)
6	T1	2.76(2.51)	1.32(1.09)	0.61(0.45)	1.37(1.14)	0.92(0.75)	3.08	1.10	1.72	1.22(1.05)	2.21(1.82)	0.94(0.76)	1.40	0.75(0.57)
7	T1	1.67(1.27)	0.71(0.54)	0.54(0.40)	1.53(1.36)	0.65(0.50)	1.34	0.62	1.11	0.91(0.72)	1.40(1.04)	0.64(0.49)	1.02	0.77(0.58)
8	T1	1.93(1.47)	0.56(0.47)	0.40(0.33)	0.64(0.55)	0.97(0.83)	3.64	0.62	1.29	1.17(1.04)	1.16(0.97)	0.98(0.86)	0.63	0.67(0.54)
9	T1	1.86(1.28)	0.49(0.41)	0.40(0.42)	0.60(0.52)	0.55(0.48)	3.30	0.40	1.12	0.87(0.76)	1.17(0.95)	0.52(0.45)	0.68	0.68(0.55)
10	T1	4.70(2.49)	3.81(2.43)	2.63(1.59)	3.65(2.27)	2.06(1.29)	1.77	3.62	3.5	2.95(1.93)	2.49(2.79)	2.02(1.26)	2.77	0.63(0.40)
11	T1	1.72(1.29)	0.78(0.51)	0.64(0.42)	0.92(0.65)	1.03(0.68)	0.41	0.92	1.68	1.15(0.79)	1.38(0.91)	1.01(0.67)	1.03	0.77(0.50)
12	T1	2.84(2.30)	1.04(0.81)	1.20(0.86)	1.25(1.02)	0.99(0.70)	0.50	1.24	1.57	1.00(0.79)	1.29(0.92)	0.89(0.64)	0.90	0.48(0.35)
13	T2	-	-	3.41(4.25)	-	3.54(4.08)	-	-	-	-	-	3.38(3.78)	6.58	0.90(1.00)
14	T2	-	-	7.29(9.80)	4.89(6.29)	9.59(12.2)	1.60	-	-	12.12(16.13)	7.91(10.3)	7.66(9.52)	7.13	0.87(1.18)
15	T2	-	-	2.73(2.21)	1.58(1.98)	2.57(3.16)	0.24	-	-	4.02(5.28)	3.65(4.73)	2.06(2.60)	1.35	0.71(0.95)
16	T3	-	-	3.18(2.11)	1.83(1.49)	2.25(1.87)	1.60	-	-	5.64(2.10)	3.90(2.52)	2.12(1.76)	2.41	0.76(0.61)
17	T3	-	-	3.01(2.52)	3.05(2.00)	3.01(1.99)	2.04	-	-	3.31(3.52)	2.44(1.97)	2.77(1.81)	4.42	1.49(1.03)
18	T3	-	-	4.46(3.23)	1.62(1.36)	2.73(2.29)	0.95	-	-	3.39(2.81)	2.14(1.57)	2.84(2.39)	2.04	0.84(0.68)
19	T3	-	-	3.58(3.98)	1.24(0.92)	1.57(1.15)	0.28	-	-	3.45(2.51)	2.60(2.86)	1.50(1.10)	3.25	0.68(0.50)
20	T2	-	-	1.94(1.66)	2.04(2.23)	2.10(2.41)	0.28	-	-	1.56(4.11)	1.86(1.44)	2.19(2.45)	2.20	1.31(1.45)
21	T3	-	-	2.56(2.02)	2.49(1.81)	3.44(2.45)	0.67	-	-	0.95(3.99)	0.85(0.99)	3.58(2.56)	3.52	0.74(0.52)
22	T3	-	-	1.16(0.92)	1.16(0.92)	1.61(1.26)	0.42	-	-	2.52(1.21)	3.06(2.54)	1.56(1.23)	1.45	0.78(0.61)
23	T2	-	-	0.66(0.79)	0.66(0.79)	0.75(0.88)	0.57	-	-	5.86(1.11)	3.38(2.32)	0.77(0.90)	0.71	0.42(0.49)
Evaluation approach		Trained w/ 12 recordings	Trained w/ 12 recordings	Trained w/ 23 recordings	Trained w/ 12 recordings	Trained w/ 23 recordings	Trained w/ 22 recordings	Trained w/ 12 recordings	Trained w/ 12 recordings	Trained w/ 22 recordings	Trained w/ 22 recordings	Trained w/ 23 recordings	Trained w/ 23 recordings	Trained w/ 23 recordings
		Results shown w/ the same 12 recordings	Results shown w/ the same 12 recordings	Results shown w/ the same 23 recordings	Results shown w/ the same 12 recordings	Results shown w/ the same 23 recordings	Results shown w/ the same 22 recordings	Results shown w/ the same 12 recordings	Results shown w/ the same 12 recordings	Results shown w/ the same 22 recordings	Results shown w/ the same 22 recordings	Results shown w/ the same 23 recordings	Results shown w/ the same 23 recordings	Results shown w/ the same 23 recordings
T1 (No. 1-12)		2.34(1.82)	1.28(1.01)	0.89(0.65)	1.50(1.12)	1.02(0.81)	1.99	1.17	1.65	1.40(1.16)	1.85(1.45)	1.02(0.81)	1.19	0.67(0.50)
T2 & T3 (No. 13-23)		-	-	3.36(3.33)	-	3.01(3.06)	-	-	-	-	-	2.77(2.74)	3.19	0.86(0.82)
T2 & T3 (No. 14-23)		-	-	3.35(3.27)	-	2.95(2.96)	0.86	-	-	4.28(4.28)	3.18(3.13)	2.71(2.63)	2.85	0.86(0.80)
All except No. 13		-	-	2.01(1.84)	-	1.90(1.98)	1.47	-	-	2.71(2.58)	2.45(2.21)	1.78(1.64)	1.94	0.75(0.64)
All (No. 1-23)		-	-	2.07(1.95)	-	1.97(1.89)	-	-	-	-	-	1.85(1.73)	2.15	0.76(0.66)

entire 23 subjects, our algorithm outperformed the existing methods.

Table 3 compares our results obtained with the BAMI-I (additional training dataset) and -II datasets (test dataset).

Because the datasets were obtained with the device developed by our team, the numerical comparisons with all the state-of-the-art methods (shown in Table 2) are not available. Instead, we performed simulations with the five very recent state-of-the-art methods: WFPV [7], FSM framework [15], Kernel-FSM framework [14], single-mode PF [11], [12], and multi-mode PF (MPF) [16]. In addition, we compared the results when only the dominant power spectrum using PPG was considered. The MPF based approach can involve various methods depending on how the estimated HR value is determined using the particles and the associated weight values. The best accuracy in the MPF results was observed when the MPF was performed with the strongest neighborhood and mean of the posterior probability densities of all particles (MPF-SN-AP). The results are summarized in the table. Our proposed model exhibited AAE and ARE of 1.39 bpm and 1.17%, respectively, with the BAMI-I dataset. These error values are lower than those of the other results—AAEs of 11.28 bpm [7], 3.88 bpm [15], 2.50 bpm [14], 6.30 bpm [11, 12], and 3.36 bpm [16]. Especially the FSM, Kernel-FSM, SPF, and MPF methods ignore low-quality signal segments or inaccurate estimation results; they consider the metric of valid HRs (VHR) as the percentage of the valid results among all the data segments based on a certain criterion such as outlier occurrence. The main purpose of the VHR is to increase the accuracy as much as possible by discarding some outlier results. The VHRs were 71.61%, 88.99%, 86.77%, and 90.91% for FSM, Kernel-FSM, SPF, and MPF, respectively. However, our method used all data segments (VHR=100%). Our proposed model also outperformed the other methods for the BAMI-II dataset. Our method exhibited AAE and ARE of 1.46 bpm and 1.23%, respectively, whereas the other methods exhibited higher AAEs: 6.09 bpm [7], 1.71 bpm [15], 2.32 bpm [14], 3.72 bpm [11], [12], and 2.90 bpm [16]. In addition, the VHRs were 72.40%, 80.74%, 84.66%, and 91.05% for FSM, Kernel-FSM, SPF, and MPF, respectively whereas that of our method was 100%.

Fig. 4 shows the estimated HR trace based on our proposed model and acceleration intensity; Figs. 4(a) and (b) show the results obtained with the ISPC dataset, whereas Figs. 4(c) and (d) show the results with the BAMI-I and BAMI-II datasets. In the top panels of Fig. 4, the estimated

TABLE 3. Comparison of the performances in terms of average absolute errors and average relative errors: BAMI-I (additional training set) and BAMI-II (test set).

Dataset	Subject	PPG only	WFPV	FSM	Kernel-FSM	SPF	MPF	Proposed Model
		AAE(ARE)	[7]	[15]	[14]	[11, 12]	[16]	AAE(ARE)
		bpm(%)	AAE(ARE)	AAE(ARE)(VHR)	AAE(ARE)(VHR)	AAE(ARE)(VHR)	AAE(ARE)(VHR)	*bpm(%)(%)*
			bpm(%)	*bpm(%)(%)*	*bpm(%)(%)*	*bpm(%)(%)*	*bpm(%)(%)*	
BAMI-I	1	46.64(37.12)	11.86(10.41)	1.18(0.92)(76.60)	1.35(1.10)(98.08)	6.40(6.78)(66.35)	1.33(1.07)(89.10)	1.06(0.85)
	2	29.46(20.71)	9.00(6.31)	1.60(1.35)(71.47)	1.65(1.33)(96.79)	12.48(13.74)(87.82)	1.65(1.32)(98.4)	1.45(1.17)
	3	17.61(15.24)	4.86(4.18)	1.35(1.22)(79.38)	1.57(1.42)(94.00)	2.09(1.91)(95.68)	1.78(1.63)(95.68)	1.34(1.2)
	4	36.59(27.32)	17.31(12.86)	15.75(11.06)(58.65)	1.15(0.94)(92.63)	33.24(39.60)(79.49)	1.18(1.00)(79.49)	0.96(0.81)
	5	39.88(33.03)	25.49(21.99)	3.70(3.07)(45.83)	2.67(2.20)(88.46)	3.43(2.92)(50.32)	3.92(3.68)(67.63)	1.88(1.57)
	6	39.29(27.09)	33.06(22.37)	33.88(23)(48.68)	12.35(8.51)(77.94)	35.48(49.11)(66.43)	13.76(15.76)(83.69)	1.39(1.16)
	7	6.83(4.67)	3.90(2.90)	1.75(1.39)(91.13)	1.83(1.43)(91.13)	2.86(2.33)(90.41)	1.98(1.54)(90.41)	1.6(1.41)
	8	9.75(7.89)	4.81(4.12)	1.91(1.66)(82.73)	2.09(1.81)(94.00)	2.52(2.21)(96.88)	2.09(1.81)(96.40)	1.77(1.55)
	9	20.42(15.62)	21.30(16.53)	1.40(1.07)(44.23)	1.75(1.34)(83.65)	2.19(1.70)(96.47)	1.72(1.33)(91.35)	1.39(1.08)
	10	33.71(22.19)	10.56(7.01)	1.65(1.1)(72.12)	1.76(1.17)(94.23)	20.79(24.06)(86.22)	22.00(25.64)(79.17)	1.34(0.89)
	11	34.01(22.57)	31.28(20.46)	1.89(1.45)(43.65)	1.84(1.31)(90.41)	2.29(1.73)(79.86)	1.91(1.38)(95.44)	1.46(1.05)
	12	17.76(13.36)	6.58(5.28)	1.56(1.45)(79.62)	1.77(1.64)(90.41)	2.19(1.97)(88.97)	1.62(1.46)(93.05)	1.31(1.17)
	13	13.13(12.12)	3.15(3.17)	1.67(1.64)(87.05)	1.66(1.61)(91.37)	1.87(1.82)(95.20)	1.67(1.62)(97.36)	1.31(1.26)
	14	5.94(5.88)	3.62(3.62)	1.66(1.76)(80.58)	1.75(1.84)(84.17)	2.62(2.81)(89.21)	2.34(2.51)(90.89)	1.2(1.25)
	15	13.51(10.14)	3.30(3.03)	3.08(2.75)(91.61)	3.08(2.75)(91.61)	2.47(2.26)(97.84)	2.91(2.61)(95.92)	3.02(2.78)
	16	20.58(14.94)	16.26(11.48)	2.23(1.65)(59.94)	2.38(1.72)(80.45)	2.45(1.78)(84.94)	2.21(1.61)(87.50)	1.53(1.13)
	17	27.29(19.66)	8.59(5.98)	1.34(1.00)(72.44)	1.38(1.00)(91.67)	1.64(1.18)(76.92)	1.50(1.09)(97.12)	1.2(0.87)
	18	27.49(17.15)	15.55(9.99)	1.51(1.11)(65.71)	1.59(1.13)(86.86)	3.17(2.33)(89.10)	5.56(4.18)(92.63)	1.07(0.77)
	19	16.28(11.85)	3.95(2.93)	1.28(1.03)(81.77)	1.42(1.12)(97.60)	1.85(1.46)(97.12)	1.29(1.02)(96.64)	1.16(0.93)
	20	15.72(11.66)	3.17(2.30)	1.48(1.13)(93.29)	1.46(1.11)(96.40)	2.42(2.02)(93.29)	1.60(1.22)(95.68)	1.24(0.93)
	21	4.77(3.22)	4.17(2.85)	1.71(1.25)(93.59)	1.74(1.26)(95.19)	2.05(1.50)(96.47)	1.77(1.29)(96.79)	1.47(1.07)
	22	8.70(7.07)	2.42(2.23)	1.47(1.34)(91.37)	1.66(1.52)(94.00)	1.85(1.68)(98.32)	1.47(1.34)(97.60)	1.23(1.11)
	23	19.67(17.73)	12.54(11.39)	1.61(1.41)(69.78)	1.73(1.54)(88.25)	4.02(4.01)(84.17)	2.40(2.69)(92.81)	1.35(1.18)
	24	7.44(7.74)	6.17(6.42)	1.41(1.53)(75.32)	1.48(1.59)(91.03)	1.88(1.99)(97.44)	1.64(1.74)(96.15)	1.04(1.11)
	25	22.82(21.52)	19.10(18.76)	8.86(9.51)(33.70)	9.44(10.25)(44.48)	3.22(3.28)(84.25)	2.64(2.57)(75.97)	0.91(0.87)
	Ave	**21.41(16.30)**	**11.28(8.74)**	**3.88(3.03)(71.61)**	**2.50(2.11)(88.99)**	**6.30(7.05)(86.77)**	**3.36(3.32)(90.91)**	**1.39(1.17)**
BAMI-II	1	4.83(3.25)	4.83(3.15)	1.29(1.05)(79.56)	1.40(1.11)(88.95)	1.68(1.28)(86.74)	1.20(0.94)(92.27)	1.30(1.03)
	2	8.99(7.69)	4.70(4.19)	1.22(1.13)(76.52)	1.33(1.21)(88.95)	1.84(1.63)(90.88)	1.22(1.10)(95.03)	1.38(1.24)
	3	1.61(1.36)	1.64(1.49)	1.65(1.42)(82.04)	1.77(1.50)(82.32)	1.74(1.45)(91.16)	1.37(1.20)(93.37)	1.16(1.03)
	4	1.99(1.49)	0.87(0.76)	1.10(0.94)(86.19)	1.18(1.01)(86.19)	1.35(1.13)(91.71)	1.04(0.88)(96.13)	0.91(0.76)
	5	10.53(8.33)	8.23(6.82)	2.05(2.00)(55.80)	2.15(2.04)(72.38)	3.09(3.07)(87.85)	16.61(19.82)(74.59)	1.53(1.49)
	6	9.38(8.78)	5.86(5.58)	1.35(1.22)(73.20)	1.52(1.41)(80.39)	1.97(1.81)(86.19)	1.43(1.29)(90.88)	1.57(1.44)
	7	5.19(4.57)	4.26(3.9)	1.60(1.34)(72.65)	1.76(1.50)(80.11)	2.01(1.70)(87.85)	1.61(1.39)(93.37)	1.40(1.21)
	8	3.36(3.29)	2.98(3.21)	2.23(2.55)(76.52)	2.31(2.62)(76.24)	3.06(3.05)(87.85)	15.07(19.33)(57.73)	1.67(1.73)
	9	17.96(12.86)	9.78(6.99)	1.49(1.21)(63.81)	1.56(1.22)(77.62)	13.95(18.02)(62.98)	1.06(0.78)(93.92)	1.42(1.05)
	10	3.11(2.89)	6.78(5.54)	1.71(1.36)(64.36)	1.98(1.54)(66.57)	2.25(1.87)(90.06)	1.86(1.60)(95.30)	1.61(1.42)
	11	16.49(13.89)	2.90(2.91)	1.84(1.88)(88.95)	1.86(1.89)(91.71)	2.30(2.09)(95.30)	1.79(1.60)(92.82)	1.53(1.36)
	12	4.96(4.61)	0.92(0.93)	1.15(0.94)(84.81)	1.18(0.97)(86.19)	1.47(1.21)(90.33)	1.25(1.05)(98.62)	1.08(0.90)
	13	3.88(2.97)	4.24(3.05)	2.02(1.80)(77.62)	2.06(1.83)(76.52)	2.22(1.92)(91.44)	1.82(1.60)(94.48)	1.39(1.17)
	14	12.46(8.26)	6.31(4.10)	1.57(1.11)(68.23)	2.21(1.59)(72.93)	1.50(0.99)(77.62)	5.98(5.34)(93.09)	1.48(0.94)
	15	14.26(10.52)	14.23(9.94)	1.90(1.60)(59.94)	2.03(1.69)(74.31)	19.03(23.29)(61.33)	2.57(2.28)(80.66)	2.63(2.33)
	16	7.17(5.03)	3.75(2.67)	1.55(1.13)(67.13)	1.60(1.16)(73.76)	1.44(0.99)(84.53)	1.04(0.73)(96.69)	1.37(1.00)
	17	4.12(3.34)	2.03(1.50)	1.30(1.00)(80.11)	1.35(1.02)(83.43)	1.44(1.10)(88.67)	1.06(0.84)(96.13)	1.20(0.95)
	18	16.48(10.31)	6.34(3.8)	1.28(0.95)(73.48)	1.33(0.96)(84.81)	1.36(0.93)(84.53)	1.10(0.77)(98.62)	1.09(0.74)
	19	11.07(7.70)	6.77(4.84)	1.19(0.95)(70.99)	1.19(0.94)(86.19)	1.56(1.18)(84.25)	1.36(1.10)(98.07)	1.06(0.84)
	20	15.88(11.46)	14.89(10.41)	2.01(1.83)(54.97)	1.84(1.67)(78.73)	2.04(1.84)(84.81)	1.33(1.14)(89.78)	1.26(1.09)
	21	10.15(7.62)	9.65(7.17)	1.63(1.28)(65.75)	2.25(1.83)(77.07)	1.29(0.93)(84.25)	1.23(0.92)(94.20)	1.79(1.34)
	22	38.94(29.12)	14.27(11.69)	3.47(3.07)(50.83)	14.89(13.65)(75.69)	14.09(19.44)(59.67)	1.13(0.85)(82.87)	1.45(1.19)
	23	10.18(9.20)	3.89(3.64)	2.62(2.31)(91.71)	2.60(2.30)(95.86)	2.96(2.69)(97.24)	2.54(2.27)(95.58)	2.34(2.11)
	Ave	**10.13(7.76)**	**6.09(4.71)**	**1.71(1.48)(72.40)**	**2.32(2.03)(80.74)**	**3.72(4.07)(84.66)**	**2.90(2.99)(91.05)**	**1.46(1.23)**

HR trace results are compared with the results when the dominant power spectrum using only PPG is considered. In addition, the bottom panels of Fig. 4 show that the acceleration intensity is associated with the HR increase. Especially, Figs. 4(a) and (b) show the results from *subjects 14* and *17*, whose data have been considered the most challenging because the measured PPG signals were severely corrupted by MAs, resulting in very low SNR. Indeed, for *subject 14* in the dataset, the reported AAEs were 6.63 bpm [1], 8.07 bpm [2], 7.29 bpm [3], 4.89 bpm [6], 9.59 bpm [7], 1.60 bpm [17], 12.12 bpm [4], 7.91 bpm [13], and 7.66 bpm [8]. For *subject 17*, the reported AAEs were 7.82 bpm [1], 7.01 bpm [2], 3.01 bpm [3], 3.05 bpm [6], 3.01 bpm [7], 2.04 bpm [17], 3.31 bpm [4], 2.44 bpm [13], and 2.77 bpm [8]. On the other hand, our proposed model provided very accurate HR estimation over the entire segments for both subjects (AAE: 0.87 bpm for *subject 14* and 1.49 bpm for *subject 17*). Figs. 4(c) and (d) also show the results when the SNR is extremely low. For *subject* 1 in the BAMI-I dataset, the AAE of 46.64 bpm was observed when only the dominant power spectrum of PPG was used. WFPV showed improved results with AAE of 11.86 bpm, which was nevertheless high. On the other hand, our results exhibited AAE of 1.06 bpm. Similarly, with *subject 22* in the BAMI-II dataset, AAE of 38.94 bpm was observed when only the dominant power spectrum of PPG was used, and

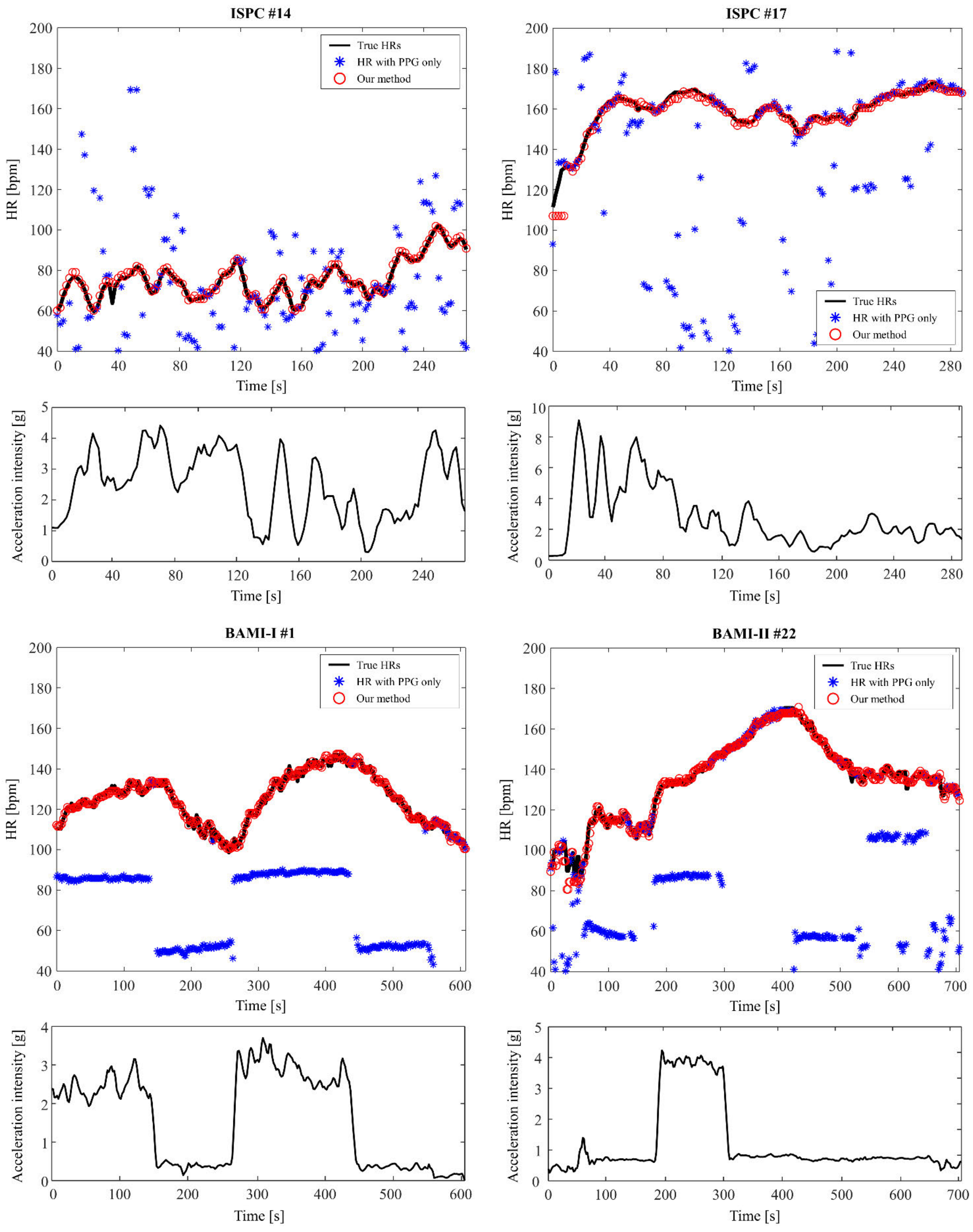


FIGURE 4. Estimated HR trace based on our proposed model (Tops) and acceleration intensity (Bottoms); (a) *Subject* 14 (ISPC), (b) *Subject* 17 (ISPC), (c) *Subject* 1 (BAMI-I) and (d) *Subject* 22 (BAMI-II).

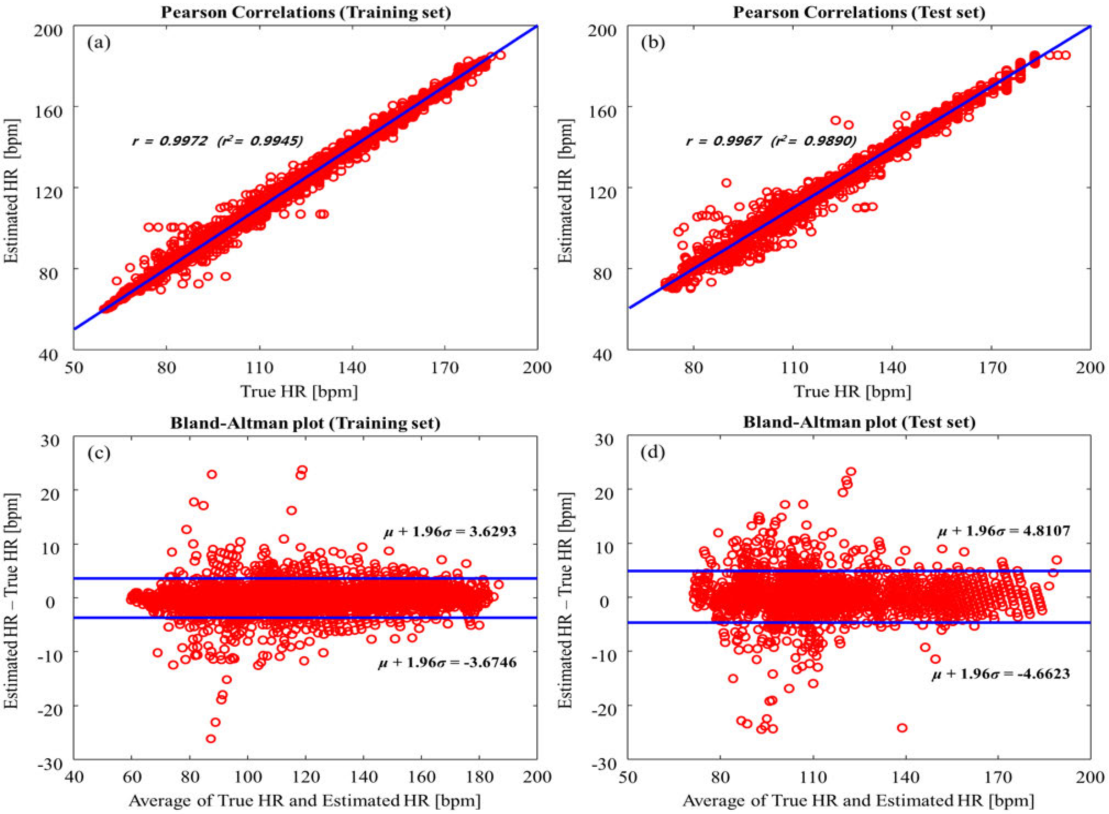


FIGURE 5. (a) Pearson correlations between estimated HRs and true HRs ($r = 0.9972$) for training dataset, (b) Pearson correlations ($r = 0.9967$) for test dataset, (c) and (d) Bland–Altman plots for training dataset ($\mu = 0.0742$ and $\sigma = 2.4166$) and test dataset ($\mu = 0.0226$ and $\sigma = 1.8632$), respectively.

WFPV exhibited AAE of 14.27 bpm. On the other hand, our results exhibited AAE of 1.45 bpm. We have presented all 71 estimation results at https://github.com/HeewonChung92/CNN_LSTM_HeartRateEstimation.

Fig. 5 shows the Pearson coefficients and Bland–Altman plots between the estimated HRs and true HRs based on the training and test datasets. As shown in Figs. 5(a) and (b), the Pearson correlation coefficients of our model were 0.9972 (r^2 =0.9945) and 0.9967 (r^2 =0.9890) for the training and test datasets, respectively.

Figs. 5(c) and (d) show the Bland–Altman plots for each dataset. The limit of agreement (LOA) for the training dataset was between –3.67 bpm and 3.63 bpm (mean 0.02 bpm, SD 1.86 bpm). Similarly, the LOA for the test dataset lay between –4.66 bpm and 4.81 bpm (mean 0.07 bpm, SD 2.42 bpm).

B. INVESTIGATION OF OPTIMIZED NETWORK

1) EFFECT OF THE TIME STEP LENGTH IN LSTM

Throughout the study, we used the time step length of six for the LSTM layers. In the case of recurrent neural networks (RNNs) such as LSTM and gated recurrent unit (GRU), the sequential information can be preserved by sharing weights over time. Thus, the long-term dependencies allow for stable tracking of the dominant frequency corresponding to the HR. However, the lengthy time steps result in large initial delay in HR estimation. In addition, the lengthy time steps may result in overfitting because the weights are shared under dynamic HR change for an extended period. Table 4 summarizes the AAE and ARE values for the training and test datasets according to the time step lengths. As summarized in Table 4, we found that the time step length of six provided the lowest AAE and ARE values. The results also indicate that overfitting occurs when the time step length exceeds six.

TABLE 4. Accuracy comparison according to the time step length.

Time step length	Training data (*n*=48) AAE (ARE)	Test data (*n*=23) AAE (ARE)
4	1.35(1.21)	1.56(1.32)
6	1.09(0.92)	1.46(1.23)
8	1.08(0.91)	1.54(1.29)
10	1.07(0.90)	1.57(1.32)
12	1.06(0.89)	1.57(1.32)

2) EFFECT OF ACCELERATION INTENSITY

Our proposed model used acceleration intensity, which was concatenated to the activations before the two LSTM layers. To investigate the effect of the acceleration intensity, we removed the acceleration intensity along with the concatenation layer and compared the accuracy. Table 5 summarizes the AAE and ARE values for the training and test datasets with and without the acceleration intensity. As summarized in Table 5, the acceleration intensity decreased the AAE and ARE values by 15.6 % and 15.8 %, respectively, for the test dataset.

TABLE 5. Accuracy comparison between with and without consideration of acceleration intensity.

	Training data (*n*=48) AAE (ARE)	Test data (*n*=23) AAE (ARE)
Our method	1.09(0.92)	1.46(1.23)
Without acceleration intensity	1.04(0.88)	1.73(1.46)

3) EFFECT OF STANDARD DEVIATION ON TRUE HR

Regarding the modified true HR probabilities, we investigated the effect of the Gaussian standard deviation by changing the value from 1 to 5. Table 6 summarizes the AAE and ARE values for the training and test datasets according to the standard deviation. As summarized in Table 6, we found that the standard deviation of three provided the lowest AAE and ARE values for the test dataset.

TABLE 6. Accuracy comparison according to the standard deviation of the Gaussian distribution on the modified true HR.

Standard deviation	Training data (*n*=48) AAE (ARE)	Test data (*n*=23) AAE (ARE)
1	1.03(0.88)	1.64(1.36)
2	1.05(0.90)	1.51(1.27)
3	1.09(0.92)	1.46(1.23)
4	1.22(1.04)	1.56(1.32)
5	1.25(1.07)	1.60(1.33)

4) COMPARISON WITH SEPARABLE CONVOLUTIONAL LAYER

In our proposed model, we aligned the power spectra from the PPG and acceleration signals with the top and bottom of the input layer; data of size 2 × 222 × 1 were fed into the 2D convolutional layer. A variant network can be constructed by realigning the two power spectra toward the depth (channel) direction, which results in input data of size 1 × 222 × 2, as shown in Fig. 6. Then, a depthwise separable convolution may be an alternative choice over a regular 2D convolution layer: the input layer is fed into the depthwise separable convolutional layer, which performs a spatial convolution independently over each channel followed by pointwise convolution (1 × 1 convolution) projecting the channels onto a new channel space [23]. For performance comparison, we modeled the variant network as shown in Fig. 6, and compared the AAE and ARE values. Table 7 summarizes the AAE and ARE values for the training and test datasets when the depthwise separable convolution layer replaces the regular 2D convolution layer in our proposed model. The AAE and ARE values increase to 1.97 bpm and 1.62%, respectively. The results indicate that the depthwise separable convolution can reduce the number of parameters (weights and biases) by 57,430, but cannot improve the overall performance. Depthwise separable convolution works efficiently when the channels are independent [23]. However, the MA components are commonly reflected in both power spectra although the PPG and acceleration signals originate from different sensor modules.

C. PERFORMANCE COMPARISON WITH ADDITIONAL DEEP LEARNING APPROACH

Apart from the CorNet [17], a few algorithms have been proposed using deep learning approach. In [24], CNN and LSTM were used for the HR estimation model. For an input layer, band-pass filtered PPG data were used, but an acceleration signal was not considered. The proposed network was trained on ISPC dataset and tested on another dataset referred to as the ADI dataset. However, it was reported that poor results were observed on the test dataset. To improve the performance, the network was additionally trained on the ADI dataset with the pre-trained model from the ISPC dataset, but the AAE was still high as 4.1 bpm. In [25], a simple fully connected layer was used for the model, where an acceleration signal was not also considered. For an input layer, 17 features from power spectrum were used. Although the AAE results were 1.39 bpm and 2.81 bpm for ISPC subject 1-12 and 13-23 datasets, respectively, the accuracy results were based on the training dataset only. Further results on test dataset (new dataset) were not provided. In [26], CNN model only was used for the HR estimation. For an input layer, power spectra of PPG and acceleration signals were used. The proposed network was separately trained on ISPC subject 1-12 and 13-23 datasets, respectively. Such session-optimized AAE values were 4.0 bpm and 16.5 bpm on ISPC subject 1-12 and 13-23 datasets, respectively. Subsequently, the network was trained on other datasets named WESAD and PPG-DaLiA with the same hyperparameters obtained from ISPC subject 13-23 datasets. The resultant AAE values were 7.47 bpm and 7.65 bpm, respectively. On the other hand, our proposed model was trained on ISPC and BAMI-I datasets, and the resultant model weights/biases were tested on BAMI-II without additional training. Furthermore, our model provided low AE values for all datasets: 0.76 bpm from ISPC, 1.39 bpm from BAMI-I and 1.46 bpm from BAMI-II.

IV. DISCUSSION AND CONCLUSION

We have presented a deep learning model for HR estimation using the power spectra from PPG and acceleration signals, and the acceleration intensity. The proposed model was sequentially structured with a 2D convolutional layer, a 1D convolutional layer, and a fully connected layer, which were incorporated for MA cancelation. It was additionally

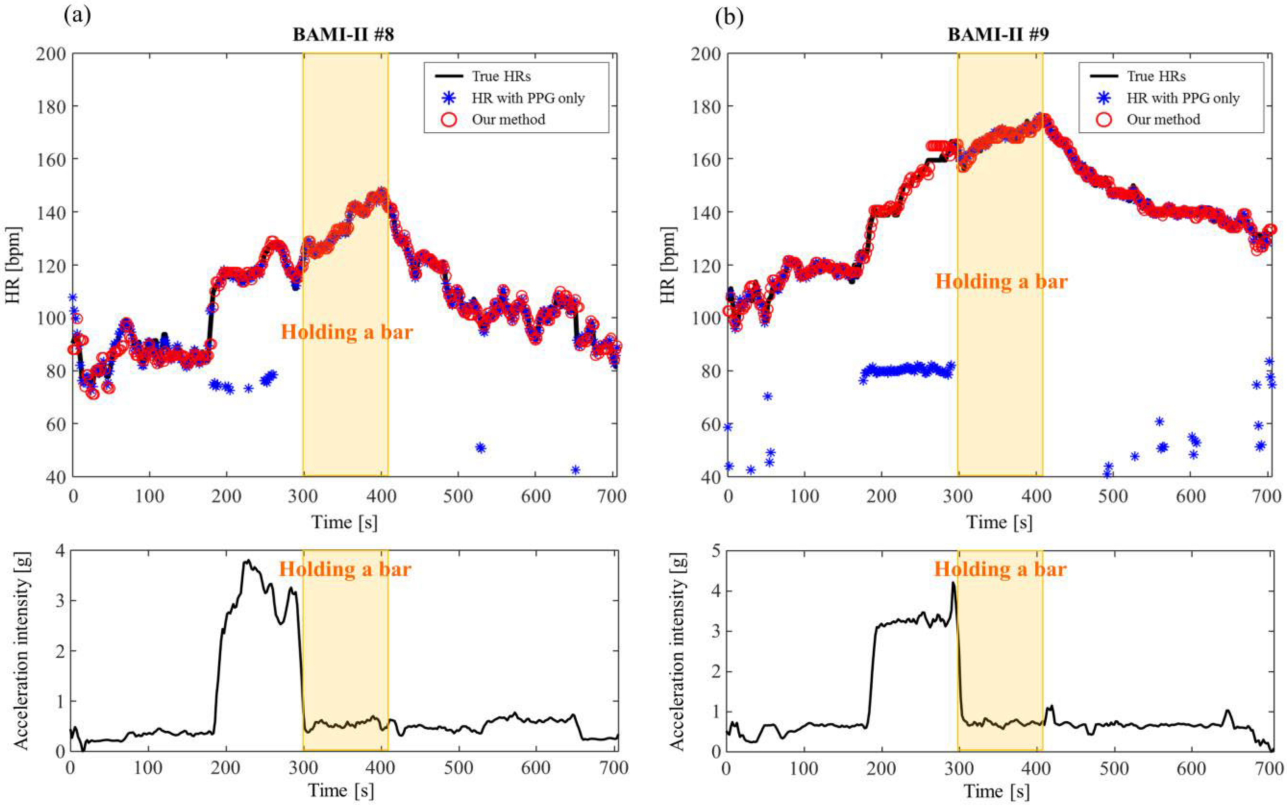


FIGURE 6. Estimated HR trace based on our proposed model (Tops) and acceleration intensity (Bottoms); (a) *Subject* 8 (BAMNI-II), (b) *Subject* 9 (BAMI-II). The exercise stage of holding treadmill bar and running is highlighted.

TABLE 7. Accuracy comparison: the two power spectra are aligned toward the depth direction, and a separable convolutional layer replaces the regular 2D convolutional layer.

	Training data (*n*=48) AAE (ARE)	Test data (*n*=23) AAE (ARE)	Total number of parameters
Our method	1.09(0.92)	1.46(1.23)	3,340,938
With Separable Convolution	2.23(1.88)	1.97(1.62)	3,283,508

structured with a concatenation layer, two LSTM layers, and a fully connected layer followed by a softmax layer, which were incorporated for HR tracking and estimation. The proposed model demonstrated AAEs of 1.09 bpm and 1.46 bpm for the training and test datasets, respectively, which exceeded the results of the current state-of-the-art methods.

To investigate the model optimization, we also considered a bidirectional LSTM layer or two GRU layers instead of the two unidirectional LSTM layers. When the two unidirectional LSTM layers were replaced by a bidirectional LSTM layer, the resultant AAEs were 0.99 bpm and 1.61 bpm for the training and test datasets, respectively. We also confirmed that further stacking of bidirectional LSTM layers did not improve the performance. When the LSTM layers were replaced by GRU layers, the resultant AAEs were 1.04 bpm and 1.64 bpm for the training and test datasets, respectively. Furthermore, we considered the regression approach to obtain the HR value rather than finding the dominant frequency bin. Based on the proposed model, we added additional fully connected layers of various depths and widths with the cost function of mean square errors, but observed that the regression approach did not learn the parameters properly. In the future, a comparative study should be conducted with an in-depth study of the network optimization of the regression approach.

Regarding the acceleration intensity information, we have shown that the acceleration intensity decreased the AAE by approximately 15%. In real life, however, the acceleration intensity may not be correlated to the exercise intensity. For instance, when a person is biking, the acceleration intensity is independent of the increase in HR. Nevertheless, in our results from BAMI-II (test dataset), the HRs were accurately estimated with low acceleration intensity even in the condition that HR increases. Fig. 6 shows the HR estimation results during the exercise stage of holding treadmill bar and running (highlighted in Fig. 6). During the exercise stage, the acceleration intensity was low, but the increased HRs were correctly estimated. It indicates that the acceleration intensity is just one of many factors for the HR estimation. The low acceleration intensity indicates that PPG signal is less corrupted by motion artifacts and the resultant power spectrum is accurately obtained, which can be interpreted that the accurate HRs can be estimated with only the accurate power spectrum of PPG signal without the acceleration intensity information. Nevertheless, in the future work, we need to investigate the real-life scenarios, especially in the case that acceleration intensity is not correlated to exercise intensity.

We also found that the estimation results were inaccurate given the simultaneous occurrence of incorrect power spectrum, no information of acceleration intensity and sudden change of HR. In Fig. 4(b), during the first 15 seconds, the dominant frequencies were around 90 or 180 bpm while the true HR was around 120bpm. In addition, the acceleration intensities were also low; and thus, the information of the acceleration intensity could not contribute to the HR estimation. Furthermore, the HR increase rate was relatively high. Such inaccurate HR estimation results were also observed in the beginning for the *subject* 22 in BAMI-II as shown in Fig. 4(d). In the future work, we will investigate the issue to minimize the inaccurate HR estimation results.

Above all, the most important aspect of future research would involve focusing on energy-efficient execution of the proposed model when implemented on wearable devices in real-time. Deep learning is undoubtedly intended to provide good performance, but the implementation on a wearable device faces many challenges because they require algorithms with low power specifications due to the limited computing power. Our proposed model includes 3,275,402 weights/biases, which require approximately 17 million multiplication and addition operations for a single HR estimation value. Thus, it is not easy to realize the real-time implementation of a deep learning model on a wearable computing platform. Reference [27] suggests that offloading deep learning workloads to the cloud is one of the feasible solutions. However, the offloading approach leads to a minimum latency of up to a few seconds, which may not be available during real-time performance. Recently, ARM announced a Compute Library (ACL), which is a comprehensive collection of low-level neural network functions optimized for the ARM Cortex CPU processors. It was reported that the processing time with ACL was reduced by 25% when compared with Tensorflow on Zuluko [27], indicating that the use of libraries optimized for deep neural network will support the realization of the proposed method on wearable devices. In addition, we can reduce the model complexity by downsizing the input data. In our proposed model, the input data, which comprised 222 frequency bin-based power spectra, had frequency resolution of 0.0122 Hz (0.73 bpm). Even if we halve the resolution to 111 frequency bins, the frequency resolution is 1.46 bpm, which is also sufficient to provide accurate HR. Furthermore, we may also consider estimating the HR by using conventional signal processing methods when the wrist movement is not large; we can estimate the HR by the proposed model only when the wrist movement is detected and the PPG signals are distorted by MAs. We hope to report on further investigation and issues related to the implementation of the proposed method in wearable devices.

REFERENCES

[1] 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 Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 522–531, Feb. 2015.

[2] Z. Zhang, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1902–1910, Aug. 2015.

[3] S. Salehizadeh, D. Dao, J. Bolkhovsky, C. Cho, Y. Mendelson, and K. 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.

[4] S. Fallet and J.-M. Vesin, "Robust heart rate estimation using wrist-type photoplethysmographic signals during physical exercise: An approach based on adaptive filtering," *Physiol. Meas.*, vol. 38, no. 2, pp. 155–170, Feb. 2017.

[5] M. Boloursaz Mashhadi, E. Asadi, M. Eskandari, S. Kiani, and F. Marvasti, "Heart rate tracking using wrist-type photoplethysmographic (PPG) signals during physical exercise with simultaneous accelerometry," *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 227–231, Feb. 2016.

[6] B. Sun and Z. Zhang, "Photoplethysmography-based heart rate monitoring using asymmetric least squares spectrum subtraction and Bayesian decision theory," *IEEE Sensors J.*, vol. 15, no. 12, pp. 7161–7168, Dec. 2015.

[7] A. Temko, "Accurate heart rate monitoring during physical exercises using PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2016–2024, Sep. 2017.

[8] M. A. Motin, C. K. Karmakar, and M. Palaniswami, "PPG derived heart rate estimation during intensive physical exercise," *IEEE Access*, vol. 7, pp. 56062–56069, 2019.

[9] E. Khan, F. Al Hossain, S. Z. Uddin, S. K. Alam, and M. K. Hasan, "A robust heart rate monitoring scheme using photoplethysmographic signals corrupted by intense motion artifacts," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 550–562, Mar. 2016.

[10] H. Lee, H. Chung, H. Ko, and J. Lee, "Wearable multichannel photoplethysmography framework for heart rate monitoring during intensive exercise," *IEEE Sensors J.*, vol. 18, no. 7, pp. 2983–2993, Apr. 2018.

[11] V. Nathan and R. Jafari, "Particle filtering and sensor fusion for robust heart rate monitoring using wearable sensors," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 6, pp. 1834–1846, Nov. 2018.

[12] Y. Fujita, M. Hiromoto, and T. Sato, "PARHELIA: Particle filter-based heart rate estimation from photoplethysmographic signals during physical exercise," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 189–198, Jan. 2018.

[13] A. Galli, C. Narduzzi, and G. Giorgi, "Measuring heart rate during physical exercise by subspace decomposition and Kalman smoothing," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1102–1110, May 2018.

[14] H. Chung, H. Lee, and J. Lee, "State-dependent Gaussian kernel-based power spectrum modification for accurate instantaneous heart rate estimation," *PLoS ONE*, vol. 14, no. 4, 2019, Art. no. e0215014.

[15] H. Chung, H. Lee, and J. Lee, "Finite state machine framework for instantaneous heart rate validation using wearable photoplethysmography during intensive exercise," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1595–1606, Jul. 2019.

[16] J. Lee, H. Chung, and H. Lee, "Multi-mode particle filtering methods for heart rate estimation from wearable photoplethysmography," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2789–2799, Oct. 2019.

[17] D. Biswas, L. Everson, M. Liu, M. Panwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Acharyya, C. Van Hoof, M. Konijnenburg, and N. Van Helleputte, "CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 2, pp. 282–291, Apr. 2019.

[18] Z. Zhang. *IEEE Signal Processing Cup 2015: Heart Rate Monitoring During Physical Exercise Using Wrist-Type Photoplethysmographic (PPG) Signals*. Accessed: Jan. 15, 2019. [Online]. Available: https://sites.google.com/site/researchbyzhang/ieeespcup2015

[19] H. Chung, H. Lee, H. Ko, and J. Lee. *CNN-LSTM Based Heart Rate Estimation From PPG and Acceleration*. Accessed: Jun. 15, 2019. [Online]. Available: https://github.com/HeewonChung92/CNN_LSTM_HeartRateEstimation

[20] S. Lu, H. Zhao, K. Ju, K. Shin, M. Lee, K. Shelley, and K. H. Chon, "Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information?" *J. Clin. Monitor. Comput.*, vol. 22, no. 1, pp. 23–29, Jan. 2008.

[21] N. Selvaraj, A. Jaryal, J. Santhosh, K. K. Deepak, and S. Anand, "Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography," *J. Med. Eng. Technol.*, vol. 32, no. 6, pp. 479–484, Jan. 2008.

[22] M. B. Mashhadi, M. Farhadi, M. Essalat, and F. Marvasti, "Low complexity heart rate measurement from wearable wrist-type photoplethysmographic sensors robust to motion artifacts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 921–924.

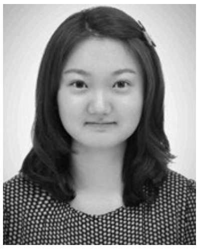
[23] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[24] A. Shyam, V. Ravichandran, S. P. Preejith, J. Joseph, and M. Sivaprakasam, "PPGnet: Deep network for device independent heart rate estimation from photoplethysmogram," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 1899–1902.

[25] M. Essalat, M. B. Mashhadi, and F. Marvasti, "Supervised heart rate tracking using wrist-type photoplethysmographic (PPG) signals during physical exercise without simultaneous acceleration signals," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Washington, DC, USA, Dec. 2016, pp. 1166–1170.

[26] 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.

[27] J. Tang, D. Sun, S. Liu, and J.-L. Gaudiot, "Enabling deep learning on IoT devices," *Computer*, vol. 50, no. 10, pp. 92–96, 2017.



HEEWON CHUNG received the dual B.S. degree in computer engineering and the M.S. degree in biomedical engineering from Wonkwang University, in 2015 and 2017, respectively. She is currently a Researcher of biomedical engineering with the Wonkwang University College of Medicine. Her current research interests include wearable computing, telemedicine systems, and biosignal/image processing.

HOON KO received the dual B.S. degree in radiology from Jeonju University, in 2015, and the M.S. degree in biomedical engineering from Wonkwang University, in 2016. He is currently a Researcher of biomedical engineering with the Wonkwang University College of Medicine. His current research interests include wearable computing, telemedicine systems, and biosignal/image processing.

HOOSEOK LEE received the dual B.S. degree in control and measurement engineering and the M.S. degree in biomedical engineering from Wonkwang University, in 2015 and 2017, respectively. He is currently a Researcher of biomedical engineering with the Wonkwang University College of Medicine. His current research interests include medical instrumentation, wearable device, and bio signal processing.

JINSEOK LEE (Senior Member, IEEE) received the dual B.S. degree in electrical engineering from Stony Brook University and Ajou University, in 2005, and the Ph.D. degree in electrical engineering from Stony Brook University, in 2009. He completed a Postdoctoral training in biomedical engineering with the Worcester Polytechnic Institute, in 2012. He is currently an Associate Professor of biomedical engineering with the Wonkwang University College of Medicine. His research interests include medical instrumentation, wearable device, bio/image signal processing, and deep neural networks.