Demographic Information Fusion Using Attentive Pooling In CNN-GRU Model For Systolic Blood Pressure Estimation

Weinan Wang[1], Pedram Mohseni[2], Kevin L. Kilgore[3] and Laleh Najafizadeh[1]

Abstract— **Fusing demographic information into deep learning models has become of interest in recent end-to-end cuff-less blood pressure (BP) estimation studies in order to achieve improved performance. Conventionally, the demographic feature vector is concatenated with the pooled embedding vector. Here, using an attention-based convolutional neural network-gated recurrent unit (CNN-GRU), we present a new approach and fuse the demographic information into the attentive pooling module. Our results demonstrate that, under calibration-based testing protocol, the proposed approach provides improved systolic blood pressure (SBP) estimation accuracy (with R^2=0.86 and mean absolute error (MAE)=4.90 mmHg) compared to both the baseline model with no demographic information fused, and the conventional approach of fusing demographic information. Our work showcases the feasibility of using attention-based methods to combine demographic features with deep learning models, and suggests new ways for fusing demographic information in deep learning models to achieve improved BP estimation accuracy.**

I. Introduction

Recently, deep learning architectures have received increased attention for developing models that estimate blood pressure (BP) from cardiovascular signals, such as the electrocardiogram (ECG) and/or the photoplethysmogram (PPG), with the goal of replacing traditional cuff-based BP measurement methods. The goal of utilizing deep learning models is to exert their data-driven end-to-end feature learning capabilities, to train models that optimally determine the information to be extracted from the physiological signals for providing accurate estimations of BP.

Differences in demographic characteristics among subjects could affect the relationship between the physiological signals and the BP values. For example, [1] shows that the same level of reduction in dicrotic notch of the arterial pulse wave (from Class I to IV) corresponds to 11.7 mmHg increment in the mean systolic blood pressure (SBP) in elderly subjects, while it only accounts for 0.9 mmHg increment in the mean SBP in younger subjects. As such, a proper BP estimation model needs to adjust how it weighs for the reduction of the dicrotic notch in the PPG signal (as an indicator of elevated BP) for subjects in different age groups.

To address this issue and to further improve BP estimation accuracy, fusing subject's demographic information (e.g., age, gender, height and weight) into deep learning models has been suggested. A conventional approach to achieve this goal is to concatenate the demographic feature vector with the embedding vector generated by the deep learning model from the input physiological signals [2]–[6], to expand the usable features for BP calculation. However, research on adversarial autoencoders [7] suggests that concatenating the class labels with the embedding generated by the encoder drives the encoder to preserve class-independent information in the learned embedding. Therefore, fusing demographic information through concatenation with the embedding vector may not allow the deep learning model to effectively map characteristics of the physiological signals to BP values for specific demographics.

To overcome the limitations of the conventional embedding concatenation method, in this study, we propose to fuse the demographic information into the attentive pooling module of the deep learning model, in order to dynamically select important temporal frames in the physiological signals that are most informative of BP under each demographic characteristic. Attentive pooling enables deep learning models to focus on frames of the input signals that are most relevant to the target estimation, by training an attention module that weighs each frame with respect to the information it carries [8]. This method has been previously adopted for the problem of BP estimation [6], [9], [10]. However, while demographic characteristics have shown to modulate the relationship between the temporal characteristics of the physiological signals (e.g., the dicrotic notch) and the BP values, prior models did not consider the demographic information as part of their attentive pooling module to determine the importance of each frame. In this study, using an attention-based convolutional neural network-gated recurrent unit (CNN-GRU) model as baseline, we show that fusing demographic information into the attentive pooling module outperforms the conventional practice of concatenating the demographic feature vector with the pooled embedding generated by GRU.

The rest of this paper is organized as follows. Section II discusses the proposed method and its differences with

This work was supported by Craig H. Neilsen Foundation Award # 598202 and the National Institutes of Health Award # 5R01EB031911.

[1] Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ, USA. {ww329, laleh.najafizadeh}@rutgers.edu. [2] Department of Electrical, Computer, and Systems Engineering, Case Western Reserve University, Cleveland, OH, USA. pxm89@case.edu. [3] Department of Physical Medicine & Rehabilitation, Case Western Reserve University and The MetroHealth System, Cleveland, OH, USA. kkilgore@metrohealth.org.

TABLE I

STATISTICAL INFORMATION OF THE SUBJECT COHORT AND THE TRAINING AND TESTING SETS USED IN THIS STUDY.

Item	Training Set	Testing Set
# Subjects	1,293	
Age (years, mean ± SD)	59.0 ± 15.0	
Gender	746 Male, 547 Female	
Height (cm, mean ± SD)	162.5 ± 9.6	
Weight (kg, mean ± SD)	60.8 ± 11.7	
BMI (kg/m^2, mean ± SD)	22.9 ± 3.4	
# Segments	465,480 360 segments per subject	51,720 40 segments per subject
SBP (mmHg, mean ± SD)	115.48 ± 18.93	115.50 ± 18.85
DBP (mmHg, mean ± SD)	62.92 ± 12.08	62.94 ± 12.07

the previous practices. Section III presents the results and discusses the advantages of the proposed method. Finally, Section IV concludes the paper.

II. METHODS

A. Dataset

PulseDB [11] is a large, cleaned dataset designed for benchmarking cuff-less BP estimation methods that uses MIMIC-III [12] and VitalDB [13] as its sources. The "Supplementary Training Subset" and the "Supplementary Calibration-Based Testing Subset" of PulseDB [11] were used as the training and testing sets in this study for benchmarking the proposed demographic information fusion method. Signals in the training and testing sets were organized as 10-s segments of synchronized ECG, PPG and arterial blood pressure (ABP) signals at 125 Hz sampling rate. Reference SBP and diastolic blood pressure (DBP) were calculated for each segment as the average of beat-to-beat SBP and DBP values retrieved from the ABP signal in each segment. Demographic information, including age (years), gender, height (cm), weight (kg) and body mass index (BMI) (kg/m^2), were retrieved.

Table I summarizes the distribution of demographic characteristics among the 1,293 subjects involved in this study, as well as the BP distribution among the segments in the training and testing sets. 360 10-s ECG and PPG segments were sampled from each subject to form the training set, while 40 different segments, not overlapping with the training segments, were sampled from each subject to form the testing set. As such, deep learning models in this study are evaluated under in-distribution, calibration-based testing protocol.

In this study, the ECG and PPG signals in each segment were used to estimate the reference SBP of each segment. The demographic feature vector to be fused into the deep learning model includes age, height and weight stored in their original numerical values, as well as the gender encoded as either 0 or 1. BMI was not included in the demographic feature vector, since it can be derived from height and weight.

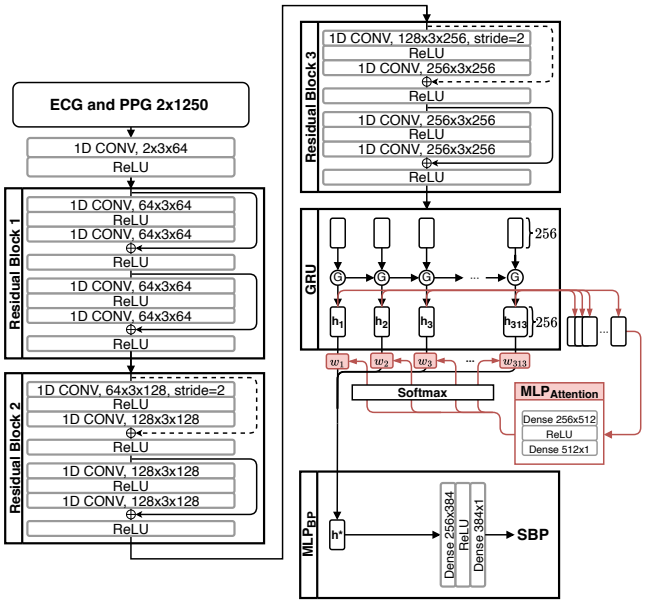


Fig. 1. The baseline **CNN-GRU** model used in this study for implementing demographic information fusion methods.

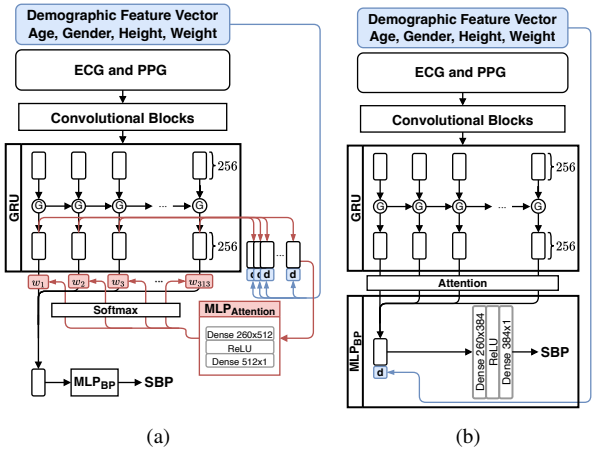
B. Attention-Based CNN-GRU Architecture

Attention-based deep learning architectures used in BP estimation studies [6], [9], [10], [14] generally consist of three components: an encoder, driven by deep learning model or manually-defined features to characterize the input physiological signals as a sequence of feature vectors; a recurrent neural network (RNN) module, which detects the sequential occurrence and changes of cardiovascular activities encoded in the feature vector sequence generated by the encoder; and an attention module, in which attentive pooling is performed to selectively collect BP-related information from the output sequence of the RNN. Specifically, let $[\mathbf{h_1}, \mathbf{h_2}, \cdots, \mathbf{h_T}]$ be a sequence of T embedding outputs of the RNN module. The objective of attentive pooling is to generate a pooled embedding vector $\mathbf{h}^*$ that gives accurate BP estimation. Overall, $\mathbf{h}^*$ is calculated as

$$\begin{aligned} \mathbf{h}^* &= \sum_{i=1}^{T} w_i \times \mathbf{h_i}, \\ w_i &= \frac{\exp(a_i)}{\sum_{j=1}^{T} \exp(a_j)}, \\ a_i &= \mathbf{f}(\mathbf{h_i}), \end{aligned} \tag{1}$$

in which $\mathbf{f}(\cdot)$ is a score evaluation function that determines the importance of each embedding vector in the sequence, which, in this study, is a feed-forward neural network with parameters learned from the data [15], [16]. The pooled feature vector $\mathbf{h}^*$ is fed into the dense layers of the model for BP calculation.

Fig. 1 depicts the baseline **CNN-GRU** model used in this study for implementing the proposed demographic information fusion method. The residual blocks in this model were based on the ResNet-18 architecture [17], with several modi-



(a) (b)

Fig. 2. Demographic information fusion methods implemented upon the baseline **CNN-GRU** model. (a) **CNN-GRU$_{\text{Attn}}$**, the proposed attentive pooling-based method. (b) **CNN-GRU$_{\text{Emb}}$**, the conventional embedding concatenation method.

fications to the number of channels, the convolutional kernel sizes, and the down-sampling factors of the model, optimized for SBP estimation when taking 10-s, 2-channel segment of ECG and PPG signals as input. The model produces a sequence of 256-dimensional feature vectors with length $T = 313$ as the GRU output. Batch normalization layers were removed from the original ResNet-18 design, as we find them to accelerate fitting on the training and the validation sets in early epochs, while degrade the generalization to the validation set in later epochs in our cases.

C. Attentive Pooling-Based Demographic Information Fusion

Fig. 2(a) describes the proposed method for fusing the demographic feature vector into the attentive pooling module of the baseline **CNN-GRU** model in Fig. 1, while Fig. 2(b) describes the implementation of the embedding concatenation method used in previous studies [4]–[6] upon the baseline **CNN-GRU** model. For the proposed attentive pooling-based method, the 4-dimensional demographic feature vector $\mathbf{d}$ consisting of age, gender, height and weight is concatenated with each of the GRU hidden states to determine the weight of each state, and the pooled embedding vector is calculated as

$$\mathbf{h}^*_{\text{Attn}} = \sum_{i=1}^{T} \frac{\exp(\mathbf{f}(\begin{bmatrix}\mathbf{h_i}\\ \mathbf{d}\end{bmatrix}))}{\sum_{j=1}^{T}\exp(\mathbf{f}(\begin{bmatrix}\mathbf{h_j}\\ \mathbf{d}\end{bmatrix}))} \times \mathbf{h_i}, \tag{2}$$

while the embedding concatenation method directly expand the pooled embedding vector $\mathbf{h}^*$ calculated in (1) as

$$\mathbf{h}^*_{\text{Emb}} = \begin{bmatrix}\mathbf{h}^*\\ \mathbf{d}\end{bmatrix}. \tag{3}$$

Regarding using either $\mathbf{h}^*$, $\mathbf{h}^*_{\text{Attn}}$ or $\mathbf{h}^*_{\text{Emb}}$ as the input of the dense layers in the model for final SBP calculation, the baseline model **CNN-GRU** (shown in Fig. 1), as well as the two other models, **CNN-GRU$_{\text{Attn}}$** (shown in Fig. 2(a)) and **CNN-GRU$_{\text{Emb}}$** (shown in Fig. 2(b)), were implemented for SBP estimation.

D. Training, Validation, and Testing Schemes

A validation set is generated by randomly sampling 10% of segments in the training set. Afterwards, all models were trained, validated and tested on the same training, validation, and testing sets, using the Adam optimizer at $1e-4$ learning rate, and the mean squared error loss. All convolutional and dense layers in each model were initialized using the Kaiming normal distribution. For each model, early-stopping was applied after 10 epochs of less than 1 mmHg^2 loss reduction on the validation set, and the weights at the end of the epoch yielding the lowest validation loss was loaded for performance evaluation on the testing set.

III. Results

Table II summarizes the SBP estimation results from the proposed attentive pooling-based demographic information fusion method, the embedding concatenation method used in previous studies [4]–[6], and the baseline model with no demographic information fused. The proposed **CNN-GRU$_{\text{Attn}}$** model yields optimal SBP estimation accuracy, with respect to achieving the highest coefficient of determination (R^2), and the lowest mean error (ME), standard deviation of error (SDE), and mean absolute error (MAE).

The embedding concatenation method implemented in **CNN-GRU$_{\text{Emb}}$** has shown inferior accuracy compared to the baseline. To investigate this performance degradation after fusing demographic information, we removed the attentive pooling module in the baseline **CNN-GRU** model, and replaced it with a simple average pooling module, namely **CNN-GRU$_{\text{Avg}}$**. Upon this model, we implemented the embedding concatenation method, namely **CNN-GRU$_{\text{AvgEmb}}$**, by using

$$\mathbf{h}^*_{\text{AvgEmb}} = \begin{bmatrix}\frac{1}{T}\sum_{i=1}^{T}\mathbf{h_i}\\ \mathbf{d}\end{bmatrix} \tag{4}$$

for SBP calculation.

The testing results from **CNN-GRU$_{\text{Avg}}$** and **CNN-GRU$_{\text{AvgEmb}}$** are summarized in Table III. **CNN-GRU$_{\text{Avg}}$** has inferior R^2, SDE and MAE compared to the baseline **CNN-GRU** model with attentive pooling module, which shows the effectiveness of attention-based embedding pooling in the **CNN-GRU** model. Meanwhile, **CNN-GRU$_{\text{AvgEmb}}$** shows improved R^2, SDE and MAE compared to **CNN-GRU$_{\text{Avg}}$** and the baseline **CNN-GRU** model, which validates the significance of fusing demographic information into deep learning models for improving the BP estimation accuracy. The performance discrepancies between **CNN-GRU$_{\text{AvgEmb}}$** and **CNN-GRU$_{\text{Emb}}$** may suggest the necessity of considering different demographic information fusion methods for different deep learning models, since some method can negatively affect the performance of the model. Nevertheless, the proposed **CNN-GRU$_{\text{Attn}}$** model remains optimal in general among all models with respect to having the highest R^2 and the lowest SDE and MAE, which stresses the feasibility

TABLE II

COMPARISON OF SBP ESTIMATION ACCURACY BETWEEN THE PROPOSED METHOD FOR FUSING DEMOGRAPHIC INFORMATION, THE CONVENTIONAL METHOD FOR FUSING DEMOGRAPHIC INFORMATION, AND THE BASELINE MODEL WITH NO DEMOGRAPHIC INFORMATION FUSED.

Model	**Calibration-based SBP Estimation**		
	R^2	**ME±SDE (mmHg)**	**MAE (mmHg)**
(Proposed) CNN-GRU$_{Attn}$	**0.86**	**0.12±7.00**	**4.90**
([4]–[6]) CNN-GRU$_{Emb}$	0.81	0.35±8.26	5.82
(Baseline) CNN-GRU	0.83	-0.65±7.81	5.40

TABLE III

TESTING RESULTS SHOWING IMPROVED SBP ESTIMATION ACCURACY VIA EMBEDDING CONCATENATION, AFTER SUBSTITUTING ATTENTIVE POOLING WITH AVERAGE POOLING IN THE BASELINE **CNN-GRU** MODEL.

Model	**Calibration-based SBP Estimation**		
	R^2	**ME±SDE (mmHg)**	**MAE (mmHg)**
CNN-GRU$_{AvgEmb}$	0.85	0.38±7.31	5.18
CNN-GRU$_{Avg}$	0.82	-0.06±8.02	5.61

of using attention-based method to combine demographic information with deep learning model.

IV. CONCLUSIONS AND FUTURE WORKS

In this study, we presented a new perspective of using attention-based methods for fusing demographic information into a deep learning model for the problem of BP estimation. Implemented on an attention-based **CNN-GRU** baseline, the proposed method achieved superior results in calibration-based SBP estimation, compared to the baseline without using demographic information, and the conventional embedding concatenation method. The performance of the embedding concatenation method is affected by including or removing attention mechanism in the baseline model that it is implemented on, which might suggest the necessity of using different demographic information fusion methods for deep learning models with different modules and structures.

This study has limitations. The BP estimation performance of the proposed method is only validated for SBP estimation, and is limited to calibration-based testing protocol in which the training and testing sets share non-overlapping data from the same subject cohort. Cuff-less BP estimation methods are expected to function well in calibration-free circumstances, that is, to estimate BP accurately in realistic situations in which the model has no data available from the testing subjects. We therefore would like to investigate the performance of attention-based demographic information fusion for both SBP and DBP estimation in future works, as well as to explore its potential for improving the inter-subject generalization capability of BP estimation models, for fulfilling practical, low-cost and reliable estimation of cuff-less BP.

REFERENCES

[1] T. R. Dawber, H. E. Thomas Jr, and P. M. McNamara, "Characteristics of the dicrotic notch of the arterial pulse wave in coronary heart disease," *Angiology*, vol. 24, no. 4, pp. 244–255, 1973.

[2] S. Shimazaki, S. Bhuiyan, H. Kawanaka, and K. Oguri, "Features extraction for cuffless blood pressure estimation by autoencoder from photoplethysmography," in *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 2857–2860.

[3] S. Shimazaki, H. Kawanaka, H. Ishikawa, K. Inoue, and K. Oguri, "Cuffless blood pressure estimation from only the waveform of photoplethysmography using CNN," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 5042–5045.

[4] D. Lee, H. Kwon, D. Son, H. Eom, C. Park, *et al.*, "Beat-to-beat continuous blood pressure estimation using bidirectional long short-term memory network," *Sensors*, vol. 21, no. 1, p. 96, 2021.

[5] S. Yang, Y. Zhang, S.-Y. Cho, R. Correia, and S. P. Morgan, "Non-invasive cuff-less blood pressure estimation using a hybrid deep learning model," *Optical and Quantum Electronics*, vol. 53, pp. 1–20, 2021.

[6] Y. Zhang, X. Ren, X. Liang, X. Ye, and C. Zhou, "A refined blood pressure estimation model based on single channel photoplethysmography," *IEEE Journal of Biomedical and Health Informatics (JBHI)*, vol. 26, no. 12, pp. 5907–5917, 2022.

[7] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," in *4th International Conference on Learning Representations (ICLR)*, 2016.

[8] F. A. Rezaur rahman Chowdhury, Q. Wang, I. L. Moreno, and L. Wan, "Attention-based models for text-dependent speaker verification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 5359–5363.

[9] H. Eom, D. Lee, S. Han, Y. S. Hariyani, Y. Lim, *et al.*, "End-to-end deep learning architecture for continuous blood pressure estimation using attention mechanism," *Sensors*, vol. 20, no. 8, p. 2338, 2020.

[10] C. El-Hajj and P. A. Kyriacou, "Deep learning models for cuffless blood pressure monitoring from PPG signals using attention mechanism," *Biomedical Signal Processing and Control*, vol. 65, p. 102301, 2021.

[11] W. Wang, P. Mohseni, K. L. Kilgore, and L. Najafizadeh, "PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods," *Frontiers in Digital Health*, vol. 4, p. 277, 2022, doi: 10.3389/fdgth.2022.1090854.

[12] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, pp. 1–9, 2016.

[13] H.-C. Lee, Y. Park, S. B. Yoon, S. M. Yang, D. Park, *et al.*, "VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients," *Scientific Data*, vol. 9, no. 1, p. 279, 2022.

[14] N. Aguirre, E. Grall-Maës, L. J. Cymberknop, and R. L. Armentano, "Blood pressure morphology assessment from photoplethysmogram and demographic information using deep learning with attention mechanism," *Sensors*, vol. 21, no. 6, p. 2167, 2021.

[15] C. Raffel and D. P. Ellis, "Feed-forward networks with attention can solve some long-term memory problems," in *4th International Conference on Learning Representations (ICLR)*, 2016.

[16] S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An attentive survey of attention models," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 12, no. 5, pp. 1–32, 2021.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.