Iterative Demographic Attentional Feature Fusion-based CNN and Transformer Network for Accurate Cuffless Blood Pressure Estimation

Liwen Tang [1], Dingchang Zheng [2] and Fei Chen [1]
[1] Southern University of Science and Technology, Shenzhen, China
E-mail: 12332129@mail.sustech.edu.cn, fchen@sustech.edu.cn
[2] Research Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom
E-mail: ad4291@coventry.ac.uk

Abstract— **Continuous cuffless blood pressure (BP) estimation is essential for cardiovascular disease monitoring. Recently, the use of deep learning models to automatically extract features and combine them with demographic features for continuous cuffless BP estimation has gained interest. Based on the observation that demographic features are highly correlated with BP estimation, this work proposes a new iterative demographic attentional feature fusion (AFF)-based CNN and Transformer network for better fusing the demographic features with the electro-cardiogram (ECG) and photoplethysmography (PPG) features, as well as accurate BP estimation. This work tested model performance using a large open BP dataset, i.e., PulseDB. The BP estimation performance of the proposed model on PluseDB dataset meets the standards of the Association for the Advancement of Medical Instrumentation (AAMI) and achieves Grade A at the British Hypertension Society (BHS) standard in the estimate of systolic blood pressure (SBP) and diastolic blood pressure (DBP). The estimations of SBP and DBP have mean absolute errors (MAE) of 3.79 mmHg and 2.37 mmHg, respectively.**

I. INTRODUCTION

Cardiovascular disease (CVD) is the most common cause of death worldwide and is now responsible for nearly 32% of all deaths, according to the World Health Organization's statistics [1]. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) are important indicators of cardiovascular health. Although the traditional cuff-based blood pressure (BP) measurement method can measure BP accurately, it is inconvenient and cannot be used to measure BP continuously. To achieve continuous and accurate BP monitor, cardiovascular signals like electrocardiogram (ECG) and photoplethysmogram (PPG) are used to estimate SBP and DBP by combining machine learning (ML) and deep learning methods.

Pulse wave analysis (PWA) is a traditional technology that estimates BP from cardiovascular signals, such as PPG. It manually extracts meaningful features from the clean signals and then finds a mapping from the extracted features and BP. Different regression ML algorithms, such as multiple linear regression [2], regression tree [3], gaussian process regression (GPR) [4], artificial neural network (ANN) [5], support vector machine (SVM), random forest (RF), and adaptive boosting (AdaBoost) [6], can be used to find the mapping between the features and BP. However, there are several drawbacks to the PWA method. First, extracting features is time-consuming [7], and another well-designed feature selection algorithm is needed for effective feature extraction [2, 4]. In addition, high-quality signals are essential during the feature extraction process.

Recently, deep learning has been used more and more in BP estimation. Unlike the ML-based PWA method, the deep learning model can automatically extract features from the row signals. Since the cardiovascular system is complex, many factors can influence BP estimation from cardiovascular signals [8]-[10]. Among these factors, demographic features (e.g., age, gender, height, and weight) are highly correlated with blood pressure, and a lot of works have tried to combine demographic features in their deep learning models [11]-[15]. Early studies only concatenate the demographic features with the features extracted by the deep learning model for later BP regression. This simple concatenation may not allow the deep learning model to effectively map characteristics of the cardiovascular signals to BP values for specific demographics. Therefore, the Attention pooling-based demographic feature fusion model is proposed to capture more essential feature frames in cardiovascular signals based on demographic features, and their works have shown that demographic features are important in BP estimation [14, 15].

On top of the research mentioned above, the aim of this study is to introduce an attentional feature fusion (AFF) module [16] to fuse features on the baseline CNN and Transformer model. Briefly, AFF will simultaneously consider the global and local characteristics of two features with inconsistent semantics and scales at the channel level and make a soft selection among these two features [16]. Meanwhile, considering a large dimensionality gap between the demographic features and cardiovascular signals features, the iterative demographic AFF-base CNN and Transformer network is proposed. In this work, the cardiovascular signal features are downsampled by convolution and average pooling layers and then fused with demographic information several times. The results show that the model proposed in this work can achieve high accuracy of BP estimation in both SBP and DBP estimation.

II. MATERIALS AND METHODS

A. Dataset

This study trains and tests the proposed model using a large dataset, PulseDB [17]. PulseDB is an open, cleaned dataset designed to benchmark BP estimation models. It uses MIMIC-III [18] and VitalDB [19] as its data sources. Synchronized ECG, PPG, and arterial blood pressure (ABP) signals were organized as 10-s segments at a 125 Hz sampling rate. The data processing details can be found in [17]. Reference SBP and

DBP were calculated for each segment as the average of beat-to-beat SBP and DBP values retrieved from the ABP signal. Demographic information, including age (years), gender, height (cm), weight (kg), and body mass index (BMI) (kg/m^2), can be obtained from the VitalDB part of PulseDB. Still, the MIMIC-III part of PulseDB can only get age information. Therefore, we only use the VitalDB part of PulseDB to train and test our model to get more usable demographic information.

PulseDB randomly splits the training and calibration-based testing subsets. These two subsets contain 1,293 subjects' ECG, PPG segments, and BP labels from the VitalDB part of PulseDB. Each subject's segments are equal in the two subsets, 360 and 40, for each subject in the training and testing subsets, respectively. In this study, deep learning models are trained and tested using these two subsets, which means the testing is calibration-based. Table I summarizes the distribution of demographic characteristics among the 1,293 subjects involved in this study and the BP distribution of the training and testing dataset.

B. *Iterative Demographic AFF-based CNN and Transformer Network*

The iterative demographic AFF-based CNN and Transformer network uses CNN and Transformer as feature extraction models to extract features from PPG and ECG signals and combines the AFF module to fuse demographic information with extracted features for better BP estimation. Figure 1 depicts the structure of the proposed model, including several modules, i.e., CNN model, Transformer encoder, AFF, iterative demographic AFF, and regression.

The CNN module is built on the ResNet architectures [20]. The input signals pass through a basic convolution layer and three residual blocks. In residual blocks, the network directly adds the input to the convolved result by shortcut connection to minimize the problem of gradient explosion or disappearance of the model during the training process. After passing through the CNN module, the original input is changed from the original two-channel signals (PPG and ECG) with 1250 sequence length in time dimension to 256-channel feature vectors with 157 sequence length in time dimension. In this process, convolutional kernels extract local features of the input within the receptive field.

TABLE I

DISTRIBUTION OF SUBJECTS' DEMOGRAPHIC CHARACTERISTICS AND BP VALUES IN TRAINING AND TESTING SUBSETS

Item	Value	
Subjects	1293	
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	
	Training subset	Test subset
Segments	465480 (360 per subject)	51720 (40 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 sequence of local feature vectors extracted by the CNN module is used as input to the Transformer encoder. The Transformer encoder computes the global correlation between the feature vectors by using a multi-head attention (MHA) mechanism and extracts the important features using weighted averaging of feature vectors. The output of MHA passes through a feed-forward layer to obtain the final sequence of feature vectors. Since the attention mechanism cannot consider the temporal order between feature vectors, the positional encoding matrix is first added to the input to introduce temporal information about the sequence. The shortcut connection and layer norm methods are also used in the Transformer module. After going through the transformer encoder, the input sequence of feature vectors is further extracted to produce a sequence of global feature vectors with the same dimension.

As local and global physiological signal features, such as morphological and statistical features, are both important for BP estimation [5], making the model consider these two kinds of features is helpful. The AFF module is used to fuse features of inconsistent semantics and scales, like global and local features. Firstly, the points-wise summation of two features with different semantics is done. Then, AFF focuses on both local and global characters in each channel of the summation result and generates a weight mask of the same dimension. Point-wise soft selection between two features is done based on the mask. Then, the output features can include information about both features with different semantics. Here, local features extracted by the CNN model and global features extracted by the Transformer encoder are fused by AFF.

Similarly, demographic information reflects the overall characteristics of a person, whereas features extracted from the PPG and ECG signals reflect the local characteristics of a person's cardiovascular system. Then, AFF is used to fuse demographic information with features extracted from PPG and ECG signals. Since AFF requires the inputs to be consistent in dimension, the demographic vector is expanded to vector sequence alone the time dimension by replicating. Then, the vector sequence goes through a 1×1 convolution to make it consistent in dimension with the PPG and ECG features. Considering that the demographic vector and the PPG and ECG features are greatly inconsistent in dimension, an iterative fusion method of the demographic vector is used. Specifically, the output of AFF is downsampled by convolution and pooling layer, and then the downsampled result is fused with demographic vectors again. The work uses three-step iterative AFF, as shown in Fig. 1.

Finally, the output, which fused demographic information and cardiovascular signal features, goes through a regression model that includes two fully connected layers and ReLU to get the SBP or DBP estimation value.

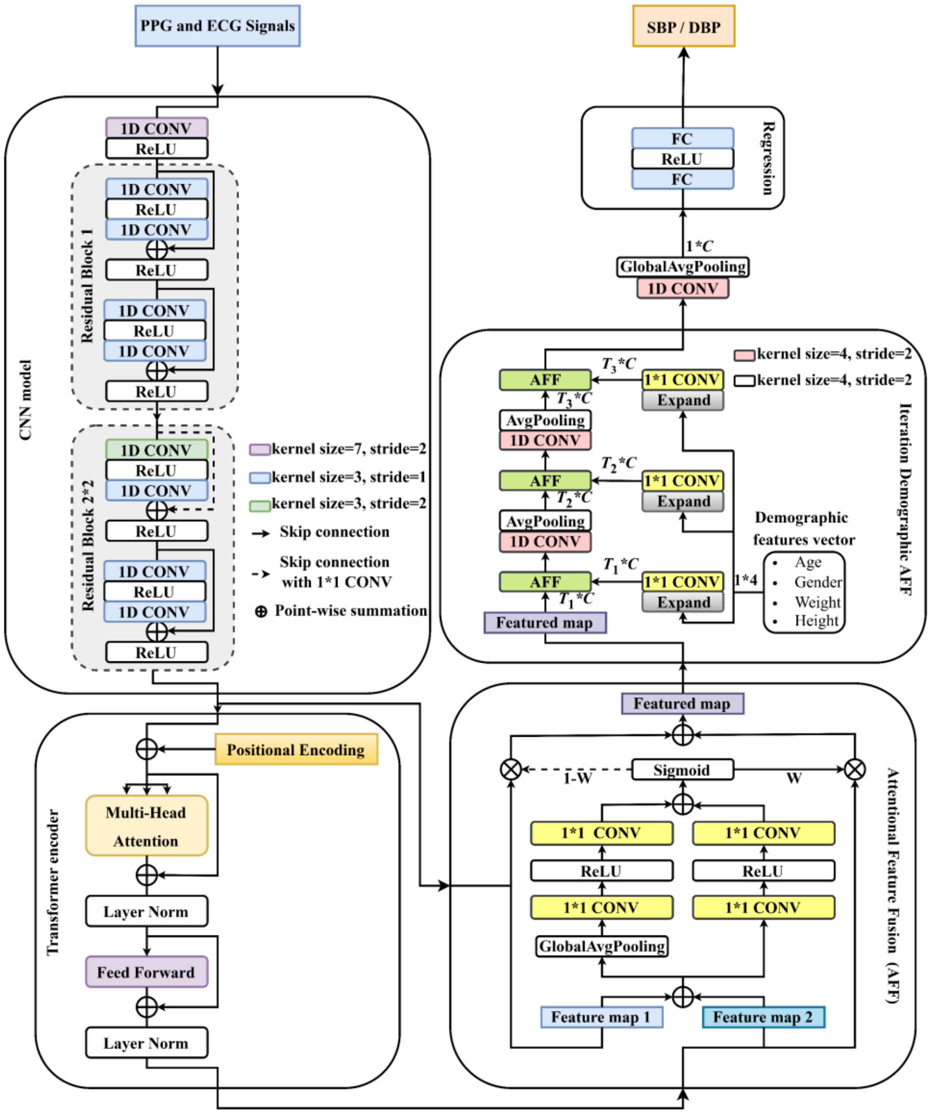

Fig. 1. Schematic diagram of iterative demographic AFF-based CNN and transformer network.

III. EXPERIMENTS

A. *Training and Testing Schemes*

The training and calibration-based testing subsets in PulseDB were used as training and testing datasets, respectively. Each PPG and ECG segment was normalized using z-score normalization. The demographic feature vector includes age, height, and weight stored in their original numerical values, as well as the gender encoded as either 0 or 1. PPG, ECG segments, and demographic feature vectors were used as input for model training and testing.

The optimization of training was selected as AdamW optimizer with 0.0001 learning rate and 128 batch size. Mean squared error (MSE) loss was selected as the model's loss function. All the model's convolutional and dense layers were initialized using the Kaiming normal distribution. Early-stopping was applied after 10 epochs of less than 0.5 loss reduction on the testing set to avoid overfitting. The model weight at the last epoch yielding the lowest loss was saved for later performance evaluation.

B. *Model Performance Evaluation Metrics and Standards*

As the BP estimation problem is a classic regression problem, we used mean absolute error (MAE) and standard derivation of error (SDE) as the primary evaluation metrics. In addition, the British Hypertension Society (BHS) and Association for the Advancement of Medical Instrumentation (AAMI) standards were used to evaluate the performance of the proposed model. The BHS standard evaluates the performance of BP estimation based on the cumulative percentages of the estimated BP with absolute errors $\leqslant$ 5 mmHg, 10 mmHg, and 15 mmHg, and it grades BP devices into different levels based on thresholds [21]. The AAMI standard requires that the MAE and SDE of the BP measurement method be $\leqslant$ 5 mmHg and 8 mmHg, respectively, and the minimum number of participants in the study is 85 [22].

C. *Ablation Experiments and Results*

To validate the effectiveness of the AFF module and iterative demographic AFF module in our proposed model, we conducted four cases of ablation experiments as follows: (1)

using only CNN and transformer encoder module, (2) using CNN, transformer encoder modules, and AFF module to fuse CNN and transformer encoder outputs, (3) using CNN, transformer encoder, AFF modules, and demographic AFF only once, (4) using CNN, transformer encoder, AFF modules, and iterative demographic AFF, which is the model proposed in this work. To keep the input of the regression model consistent dimensionality, global average pooling was performed along the sequence of feature vectors in the first three models before making BP regression.

As shown in Table II, the results of ablation experiments are consistent in SBP and DBP estimation. Model 4, which includes AFF and iterative demographic modules, achieves the minimum MAEs and SDEs, 3.79±5.33 mmHg and 2.37±3.55 mmHg for SBP and DBP, respectively. Specifically, Model 2 uses the AFF module based on Model 1 to perform feature fusion between the CNN output and the Transformer encoder output. The results show significant improvement in SBP and less in DBP, indicating that with the AFF module, the model can catch features with different scales, thus achieving better BP estimation. Model 3 further incorporates demographic features AFF only once, based on Model 2. The results show an improvement in the estimation of both SBP and DBP, which suggests that incorporating demographic features through the AFF can better predict BP. Model 4 uses iterative demographic AFF with three times iteration, achieving a better performance than Model 3. This suggests a strong correlation between demographic information and BP, and the iterative fusion method enables the model to better focus on demographic information and more accurately estimate blood pressure.

D. BP Estimation Performance Evaluation Using the BHS and AAMI Standards

Table III displays the specific grade split by BHS and the performance of the proposed model in the estimation of SBP and DBP. The cumulative percentages with absolute errors ⩽ 5 mmHg, 10 mmHg, and 15 mmHg are 74.2%, 94.3%, 98.4% in SPB estimation, 90.0%, 98.7%, and 99.7% in DBP estimation. The results show that the performance of the proposed model in SBP and DBP achieves Grade A at the BHS standard.

As shown in Table IV, The MAE and SDE of the proposed model in SBP estimation are 3.79 mmHg and 5.33 mmHg. The MAE and SDE in DBP estimation are 2.37 mmHg and 3.55 mmHg. Both SBP and DBP estimations meet the AAMI standard.

TABLE II

BP ESTIMATION PERFORMANCE WITH DIFFERENT MODULE COMBINATIONS

Model		MAE±SDE (mmHg) SBP	MAE±SDE (mmHg) DBP
1	CNN, Transformer	5.19±7.48	3.06±4.51
2	CNN, Transformer and AFF	4.60±6.70	3.03±4.53
3	CNN, Transformer, AFF, and demographic AFF once	4.48±6.35	2.62±3.51
4	CNN, Transformer, AFF, and iterative demographic AFF	**3.79±5.33**	**2.37±3.55**

TABLE III

EVALUATION OF BHS STANDARD

		Cumulative Error Percentage ⩽5 mmHg	⩽10 mmHg	⩽5 mmHg
Proposed model	**SBP**	74.2%	94.3%	98.4%
	DBP	90.0%	98.7%	99.7%
BHS	**Grade A**	60%	85%	95%
	Grade B	50%	75%	90%
	Grade C	40%	60%	85%

TABLE IV

EVALUATION OF AAMI STANDARD

		MAE (mmHg)	SDE (mmHg)	Number of subjects
Proposed model	**SBP**	3.79	5.33	1293
	DBP	2.37	3.55	
AAMI standard		⩽5	⩽8	85

IV. CONCLUSION

Based on the observation that demographic information is highly correlated with blood pressure, this work proposes an iterative demographic AFF-based CNN and transformer network. Through ablation experiments, we demonstrated that the AFF module can fuse features with different scales from the CNN module and transformer encoder module. The iterative demographic AFF module uses multi-times feature fusion to make the model fuse demographic features better. By combining both modules, the model can significantly improve BP estimation.

In addition, the evaluation results achieve Grade A at the BHS standard and meet the AAMI standard on SBP and DBP estimation. This indicates that the proposed model is highly reliable and accurate in BP estimation.

V. ACKNOWLEDGMENT

This work was supported by Shenzhen Key Technology Program Funding (JSGG20220831103803006).

REFERENCES

[1] N. Townsend, L. Wilson, P. Bhatnagar, K. Wickramasinghe, M. Rayner, and M. Nichols, "Cardiovascular disease in Europe: epidemiological update 2016," Eur. Heart J., vol. 37, no. 42, pp. 3232–3245, 2016.

[2] W.-H. Lin, F. Chen, Y. Geng, N. Ji, P. Fang, and G. Li, "Towards accurate estimation of cuffless and continuous blood pressure using multi-order derivative and multivariate photoplethysmogram features," Biomed. Signal Process. Control, vol. 63, p. 102198, Jan. 2021.

[3] S. G. Khalid, H. Liu, T. Zia, J. Zhang, F. Chen, and D. Zheng, "Cuffless blood pressure estimation using single channel photoplethysmography: a two-step method," IEEE Access, vol. 8, pp. 58146–58154, 2020.

[4] M. H. Chowdhury et al., "Estimating Blood Pressure from the Photoplethysmogram Signal and Demographic Features Using Machine Learning Techniques," Sensors, vol. 20, no. 11, p. 3127, Jun. 2020.

[5] P. Yao et al., "Multi-Dimensional Feature Combination Method for Continuous Blood Pressure Measurement Based on Wrist PPG Sensor," IEEE J. Biomed. Health Inform., vol. 26, no. 8, pp. 3708–3719, Aug. 2022.

[6] G. Thambiraj, U. Gandhi, U. Mangalanathan, V. J. M. Jose, and M. Anand, "Investigation on the effect of Womersley number, ECG and PPG features for cuff less blood pressure estimation using machine learning," Biomed. Signal Process. Control, vol. 60, p. 101942, 2020.

[7] W.-H. Lin, D. Zheng, G. Li, H. Zhou, and F. Chen, "Investigation on Pulse Wave Forward Peak Detection and Its Applications in Cardiovascular Health," IEEE Trans. Biomed. Eng., vol. 69, no. 2, pp. 700–709, Feb. 2022.

[8] M. E. Safar, J. Blacher, and P. Jankowski, "Arterial stiffness, pulse pressure, and cardiovascular disease—Is it possible to break the vicious circle?," Atherosclerosis, vol. 218, no. 2, pp. 263–271, Oct. 2011.

[9] W.-H. Lin, D. Zheng, G. Li, and F. Chen, "Age-Related Changes in Blood Volume Pulse Wave at Fingers and Ears," IEEE J. Biomed. Health Inform., pp. 1–11, 2024.

[10] W.-H. Lin, X. Li, Y. Li, G. Li, and F. Chen, "Investigating the physiological mechanisms of the photoplethysmogram features for blood pressure estimation," Physiol. Meas., vol. 41, no. 4, p. 044003, May 2020.

[11] Y. Zhang, X. Ren, X. Liang, X. Ye, and C. Zhou, "A Refined Blood Pressure Estimation Model Based on Single Channel Photoplethysmography," IEEE J. Biomed. Health Inform., vol. 26, no. 12, pp. 5907–5917, Dec. 2022.

[12] D. Lee et al., "Beat-to-Beat Continuous Blood Pressure Estimation Using Bidirectional Long Short-Term Memory Network," Sensors, vol. 21, no. 1, p. 96, Dec. 202.

[13] 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," Opt. Quantum Electron., vol. 53, no. 2, p. 93, Feb. 2021.

[14] W. Wang, P. Mohseni, K. L. Kilgore, and L. Najafizadeh, "Demographic Information Fusion Using Attentive Pooling In CNN-GRU Model For Systolic Blood Pressure Estimation," in 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia: IEEE, Jul. 2023, pp. 1–4.

[15] Z.-D. Liu et al., "HGCTNet: Handcrafted Feature-Guided CNN and Transformer Network for Wearable Cuffless Blood Pressure Measurement," IEEE J. Biomed. Health Inform., vol. 28, no. 7, pp. 3882–3894, Jul. 2024.

[16] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional Feature Fusion," in 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA: IEEE, Jan. 2021, pp. 3559–3568.

[17] 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," Front. Digit. Health, vol. 4, p. 1090854, Feb. 2023.

[18] A. E. W. Johnson et al., "MIMIC-III, a freely accessible critical care database," Sci. Data, vol. 3, no. 1, p. 160035, May 2016.

[19] H.-C. Lee, Y. Park, S. B. Yoon, S. M. Yang, D. Park, and C.-W. Jung, "VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients," Sci. Data, vol. 9, no. 1, p. 279, Jun. 2022.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778. Accessed: Jul. 15, 2024.

[21] E. O'Brien et al., "The British Hypertension Society protocol for the evaluation of blood pressure measuring devices," J Hypertens, vol. 11, no. Suppl 2, pp. S43–S62, 1993.

[22] G. S. Stergiou et al., "A Universal Standard for the Validation of Blood Pressure Measuring Devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement," Hypertension, vol. 71, no. 3, pp. 368–374, Mar. 2018.