2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

Comparison of Nine Deep Regressors in Continuous Blood Pressure Estimation Using Single-Channel Photoplethysmograms under the PulseDB

Takumi Yamamoto[1,2], Suguru Kanoga[1], and Yuta Sugiura[2]

Abstract— **Blood pressure (BP) is a vital parameter in medical treatment and diagnosis, and as a non-invasive method to measure BP, some deep learning models have been proposed to estimate BP from photoplethysmograms (PPGs). However, the datasets and the method of dividing them into training and testing subsets are not uniform, making it difficult to compare them fairly. In this study, we compared the performance of nine deep learning models for estimating systolic and diastolic BP using PPGs. We used PulseDB, which has "calibration-based subset" and "calibration-free subset" as test subsets. The calibration-based subset has the same subject's data in the training subset, and the calibration-free subset does not have the same subject's data in the training subset. The results showed that ST-ResNet performed the best, and it is important to evaluate the models using both calibration-based and calibration-free subsets, and to prevent the overfitting by using the weight decay.**

I. INTRODUCTION

Blood pressure (BP) stands as a pivotal parameter in both the diagnosis and treatment of cardiovascular disease (CVD) [1]. In 2020, an estimated 19.05 million deaths globally were attributed to CVDs, marking an increase of 18.71% from 2010 [2]. Continuous monitoring of BP is indispensable for the prevention of CVDs. Specifically, the assessment of systolic (SBP) and diastolic blood pressure (DBP), representing maximum and minimum BP, is crucial in this monitoring process.

BP measurements have been conducted using cuff-based devices [3]. The utilization of such devices necessitates strict adherence to a prescribed measurement protocol [4] and typically relies on access to healthcare facilities such as hospitals and clinics. Moreover, the presence of stress or anxiety among certain patients during measurements has been identified as a potential source of inaccuracies [5]. Thus, there exists an increasing need for measurement techniques that offer greater accessibility and minimize user stress.

Many researchers have proposed cuff-less BP measurement/estimation techniques using wearable devices. Recently, commercially available devices like the Apple Watch, equipped with photoplethysmogram (PPG) sensors, have garnered attention in this domain. With the rapid advancement of machine learning, deep learning-based models [6], [7], [8], [9] have become mainstream, although some shallow algorithms like MLP [10], support vector regression [11], and random forest [12] have also been proposed.

When evaluating such models, comparing the proposed model's performance to others isn't fair across papers. Dataset variations and different methods of splitting training and testing subsets contribute to this disparity. There are two ways to divide these subsets: including the same subjects in both [6], [7], [13] or different subjects [8]. While performance tends to be higher when subjects are the same, achieving high performance with different subjects is crucial for practical applications. Thus, evaluating model performance should involve a unified dataset and consideration of both conditions in subset division.

In this study, we compared the performance of deep-learning models in estimating SBP and DBP using only PPG signals, employing a unified dataset. We utilized PulseDB [14], published in 2023, involving two types of datasets: a calibration-based subset with same subjects in both training and testing subsets, and a calibration-free dataset with distinct subjects. These subsets allow us to use the unified dataset and study both conditions with and without the same subjects as the training data.

Our research questions are as follows:

- RQ1: Is there a difference of BP estimation performance between calibration-based and calibration-free conditions?
- RQ2: Which deep learning models perform best under different conditions?

II. MATERIALS AND METHODS

A. Dataset and Preprocessing

In this study, we utilized the PulseDB dataset [14] containing 5,361 subjects, which is based on the MIMIC-III and VitalDB databases (2,423 subjects from MIMIC-III and 2,938 subjects from VitalDB). PPG data were sampled at a rate of 125 Hz and segmented into 10-second windows. As a result of the segmentation and cleaning process, 5,245,454 windows are included in this dataset. The windows were filtered with a fourth order Chebychev-II filter at [0.5, 8] Hz. In addition, the first derivative of PPG (velocity PPG: VPG) and second derivative of PPG (acceleration PPG: APG) were calculated to add information to the single-channel PPG. In order to align their lengths with that of the PPG

*This work was partially supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B), Grant Number JP23H03445 and JP23K28135.

[1]T. Yamamoto and S. Kanoga are with the National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan.

[2]T. Yamamoto and Y. Sugiura are with Department of Information and Computer Science, Faculty of Science and Technology, Keio University, Kanagawa, Japan imuka06x17@keio.jp.

data, the initial samples in the VPG and APG were padded with their respective values. Previous studies have found that downsampling by about a quarter does not significantly degrade performance compared to the original state (see Table II in [14]). We downsampled the PPG, VPG, and APG data from 1250 to 250 samples to reduce the computational complexity. Each window was normalized to zero mean and unit variance.

This dataset provides training and testing subsets that divide the total data in various ways, as described in Table 4 of the previous study [14]. We utilized the training, calibration-based testing, and calibration-free testing subsets. The calibration-based testing subset comprises an additional collection of window data sourced from the identical subject as the training subset. Conversely, the calibration-free testing subset exclusively contains window data from subjects distinct from those in the training subset. Thus, the calibration-free testing subset poses a more challenging environment for BP estimation compared to the calibration-based testing subset. In this study, 10% of the data randomly selected from the training subset was used as the validation subset.

B. Architectures

To fairly compare the performance of deep-learning models in SBP and DBP estimation, we selected nine architectures: (a) PP-Net [6], (b) mPP-Net [7], (c) BP-CRNN [15], (d) modified-LRCN [16], (e) IMSF-Net [17], (f) 1D ResNet18 [18], (g) the model used by Wang et al. [19] (which we named "Attention-Based CNN-GRU"), (h) the model proposed by Emo et al. [20] (which we named "DeepCNN-BiGRU-Attention"), and (i) ST-ResNet [8]. Fig. 1 shows the architecture overview. We decided to keep hyperparameters such as kernel size the same as in the previous studies.

PP-Net [6] consists of two convolutional layers, two long short-term memory (LSTM) layers, and one dense layer (Fig. 1(a)). mPP-Net is modified version of PP-Net [7] and has one additional convolutional layer (Fig. 1(b)). Note that we have modified it from their original version by adding the batch normalization layer. BP-CRNN [15] consists of three convolutional layers, one gated recurrent unit (GRU) layer, and two dense layers (Fig. 1(c)). There is skip connection from the first output to third output of convolutional layers. modified-LRCN [16] consists of two multi-convolution modules with different kernel sizes, one convolutional layer, two LSTM layers, and one dense layer (Fig. 1(d)). IMSF-Net [17] consists of two IMSF blocks and efficient channel attention (ECA), one convolutional layer, one LSTM layer, and one dense layer (Fig. 1(e)). An IMSF block conducts feature extraction in a two-channel way using different kernel size convolutional layers. In addition, ECA is a kind of local cross-channel interaction strategy without dimensionality reduction [21]. 1D ResNet18 [18] is a 1D version of ResNet18, which consists of four Res blocks (Fig. 1(f)). The sizes of output channel from the first, second, third, and fourth Res block's convolutional layers were 64, 128, 256, and 512, respectively. Attention-Based CNN-GRU [19] consists of three Res blocks, one GRU layer with attention mechanism, and two dense layers (Fig. 1(g)). The sizes of output channel from the first, second, or thirt Res block's convolutional layers were 64, 128, and 256, respectively. DeepCNN-BiGRU-Attention [20] consists of a VGGNet-based convotional neural network structure, one bidirectional GRU (BiGRU) layer with a feed-forward attention mechanism, and one dense layer (Fig. 1(h)). ST-ResNet [8] consists of spectral-temporal (ST) blocks, Res blocks, one GRU layer, and three dense layers (Fig. 1(i)). Each spectrogram for PPG, VPG, and APG data were calculated based on short-time Fourier transform.

C. Implementation, Training, and Evaluation

The deep-learning models described in Section II-B were implemented through Python 3.11.6 and PyTorch 2.0.1. It was performed on a 64-bit Windows 11 machine equipped with an Intel Xeon Gold 6330 CPU and 256 GB of RAM.

These models were trained on an NVIDIA GeForce RTX A6000 GPU and the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$). We used the mean squared error (MSE) as the loss function, and as a learning rate scheduler, `ReduceLROnPlateau` scheduler was applied to all models. The initial learning rate, the number of patience, batch size, the number of epochs were set to 0.001, 5, 300, and 10. The more parameters in the model, the easier it is to overfitting, which is advantageous for calibration-based models, but also because regularization may provide general features for calibration-free models. To check the model training performance from multiple perspective, we prepared two conditions: (1) weight decay was 0.001 and (2) weight decay was 0.

For evaluation, we calculated the mean absolute error (MAE) between true and estimated SBP and DBP values for the nine deep-learning models. In addition, the BP estimation performance was checked based on the British Hypertension Society (BHS) standard [22].

III. Results and Discussion

A. Model Performance

Table I presents MAEs comparing true and estimated SBP and DBP values for the nine deep-learning models. These comparisons were conducted across calibration-based and calibration-free testing subsets, utilizing two types of weight decay values. Consistent with previous findings [23], all models had more difficulty estimating SBP values than DBP values. The inspection for calibration-free testing subset is said to be more difficult than the one for calibration-based testing subset [23], [14], and approximately this trend was observed. However, this trend was slightly different with and without weight decay. In the case of the calibration-free testing subset, MAE tended to be larger without weight decay than with it. Conversely, in the case of the calibration-based testing subset, MAE tended to be smaller than in the case of the calibration-free testing subset without weight decay. This can be interpreted as a state in which the calibration-based testing subset can overfit the training subset because it has the same subject data as the training subset. Inherently, the deep-learning model cannot feed the entire

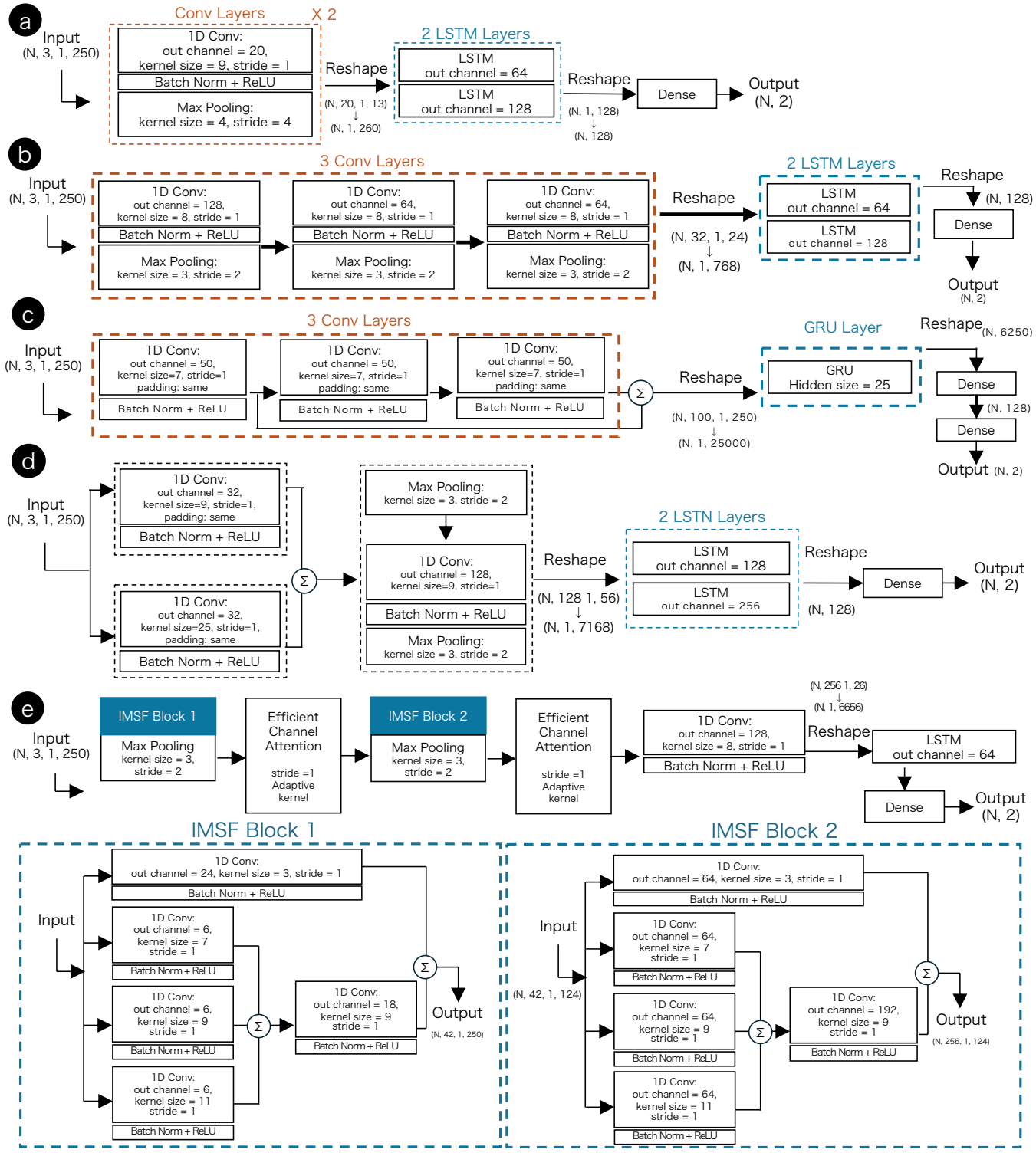

Fig. 1. Architecture overview of (a) PP-Net, (b) mPP-Net, (c) BP-CRNN, (d) modified LRCN, (e) IMSF-Net.

human population with data, and the majority of users are not included in the training subset. Therefore, it is important to have high accuracy with respect to the calibration-free testing subset, and in this sense, the inclusion of weight decay would be useful for acquiring general-purpose features. In the condition with weight decay, ST-ResNet demonstrated the best performance. In the condition without weight decay, ST-ResNet demonstrated the best performance within the calibration-based subset, while m-PPNet stood out in the calibration-free subsets for SBP values, and BP-CRNN for DBP values. The performance of these models is thought to be due to the relatively small architectural scale (i.e., small

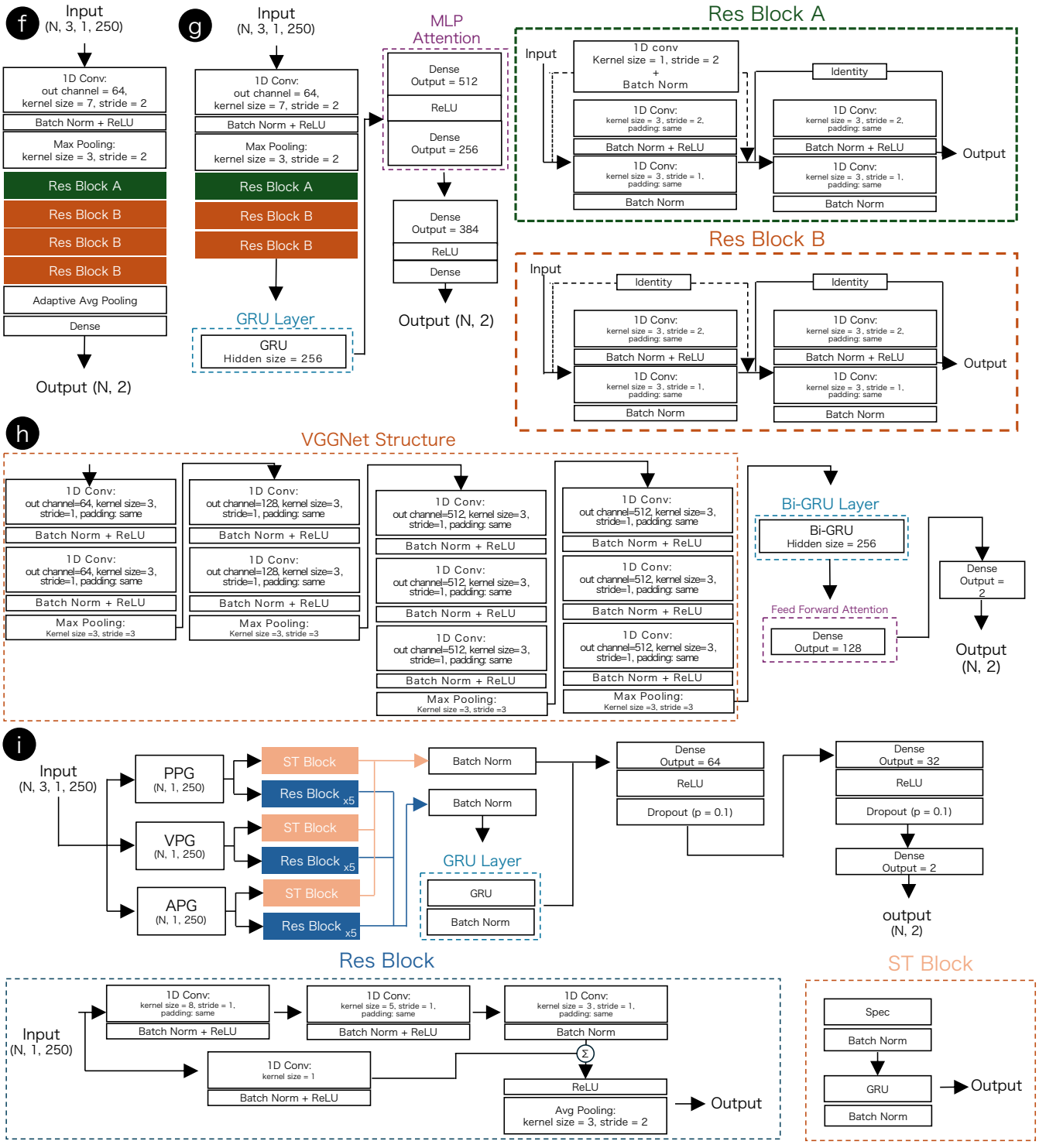

Fig. 1. Cont. (f) 1D ResNet18, (g) Attention-Based CNN-GRU, (h) DeepCNN-BiGRU-Attention, (i) ST-ResNet.

number of parameters), which makes it difficult to overfit.

Tables II and III show the results of comparative analysis with BHS standard for the calibration-based and calibration-free testing subsets. The BHS standard categorizes models into four grades (i.e., A, B, C, and not applicable). In the condition with weight decay, all models did not meet Grade C in both SBP and DBP values. In the condition without weight decay, IMSF-Net, 1D ResNet18, Attention-Based CNN-GRU, and DeepCNNGRUAttention achieved Grade C for DBP values in the calibration-based subset, while ST-ResNet achieved Grade B for DBP in the calibration-based subset. other models did not meet Grade C. As for SBP, none

TABLE I

MAEs COMPARING TRUE AND ESTIMATED SBP AND DBP VALUES FOR THE NINE DEEP-LEARNING MODELS. THE VALUES IN BOLD ARE THE MOST ACCURATE RESULTS FOR THE DEEP-LEARNING MODEL IN EACH CONDITION AND TARGET.

Model	Calibration-based				Calibration-free			
	w weight decay		w/o weight decay		w weight decay		w/o weight decay	
	SBP	DBP	SBP	DBP	SBP	DBP	SBP	DBP
PP-Net	15.01	9.00	13.80	8.28	14.75	8.94	14.09	8.56
mPP-Net	14.98	8.91	13.33	8.00	14.74	8.87	**14.01**	8.58
BP-CRNN	14.28	8.57	13.61	8.18	14.29	8.54	14.14	**8.55**
m-LRCN	14.70	8.84	13.53	8.20	14.44	8.84	14.74	9.14
IMSF-Net	13.61	8.16	12.38	7.39	13.85	8.45	15.46	9.55
1D ResNet18	13.59	8.18	12.77	7.72	13.94	8.45	15.98	9.85
Attention-Based CNN-GRU	14.53	8.68	12.50	7.49	14.37	8.70	15.31	9.51
DeepCNN-BiGRU-Attention	13.60	8.18	11.76	7.09	13.85	**8.43**	15.99	9.76
ST-ResNet	**13.52**	**8.09**	**10.67**	**6.37**	**13.84**	**8.43**	15.32	9.59

TABLE II

COMPARISON ANALYSIS WITH THE BHS STANDARD FOR CALIBRATION-BASED TESTING SUBSET BASED ON CUMULATIVE ERROR PERCENTAGE.

Model	w weight decay								w/o weight decay							
	SBP				DBP				SBP				DBP			
	$\leqq 5$	$\leqq 10$	$\leqq 15$	Grade	$\leqq 5$	$\leqq 10$	$\leqq 15$	Grade	$\leqq 5$	$\leqq 10$	$\leqq 15$	Grade	$\leqq 5$	$\leqq 10$	$\leqq 15$	Grade
PP-Net	21.19	40.80	58.18	–	34.80	63.54	82.46	–	23.35	44.67	62.52	–	38.10	67.79	85.85	–
mPP-Net	21.23	40.83	58.38	–	35.13	64.03	82.88	–	24.22	46.30	64.27	–	39.42	69.52	87.01	–
BP-CRNN	22.55	43.47	60.78	–	36.48	65.86	84.60	–	23.83	45.55	63.21	–	38.55	68.34	86.14	–
m-LRCN	21.96	42.02	59.47	–	38.41	64.24	83.21	–	24.58	46.28	63.99	–	39.41	68.65	85.62	–
IMSF-Net	23.73	45.48	63.18	–	38.52	68.50	86.36	–	28.18	51.67	68.84	–	45.36	73.56	88.34	C
1D ResNet18	23.88	45.52	63.16	–	35.46	68.21	86.26	–	27.23	50.22	67.26	–	43.26	71.75	87.12	C
Attention-Based CNN-GRU	21.96	42.56	60.13	–	35.82	64.97	84.06	–	27.31	50.86	68.39	–	44.39	73.26	88.03	C
DeepCNN-BiGRU-Attention	23.93	45.56	63.24	–	38.28	68.30	86.21	–	30.73	54.85	71.37	–	48.08	75.60	88.99	C
STResNet	23.86	45.82	63.54	–	38.93	68.95	86.61	–	33.26	58.86	75.57	–	52.19	79.79	91.62	B
BHS	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A
	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B
	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C

TABLE III

COMPARISON ANALYSIS WITH THE BHS STANDARD FOR CALIBRATION-FREE TESTING SUBSET BASED ON CUMULATIVE ERROR PERCENTAGE.

Model	w weight decay								w/o weight decay							
	SBP				DBP				SBP				DBP			
	$\leqq 5$	$\leqq 10$	$\leqq 15$	Grade	$\leqq 5$	$\leqq 10$	$\leqq 15$	Grade	$\leqq 5$	$\leqq 10$	$\leqq 15$	Grade	$\leqq 5$	$\leqq 10$	$\leqq 15$	Grade
PP-Net	21.18	40.58	58.03	–	33.88	63.48	82.70	–	21.92	42.90	60.90	–	35.70	65.70	84.27	–
mPP-Net	21.23	40.66	58.07	–	34.16	65.47	82.83	–	22.62	43.51	61.49	–	36.09	65.42	84.09	–
BP-CRNN	22.05	42.29	60.22	–	36.62	65.62	84.17	–	22.47	43.28	60.78	–	36.54	65.78	84.17	–
m-LRCN	21.70	41.68	59.34	–	34.37	64.13	83.33	–	21.94	42.34	59.55	–	34.67	62.93	81.21	–
IMSF-Net	22.42	43.89	62.08	–	37.04	66.10	84.45	–	21.34	40.85	57.42	–	33.15	60.49	79.44	–
1D ResNet18	22.46	43.57	61.08	–	36.92	66.18	84.26	–	20.52	39.73	56.21	–	32.42	59.24	78.04	–
Attention-Based CNN-GRU	22.00	42.21	59.72	–	35.53	64.97	83.60	–	21.40	41.27	58.16	–	33.54	60.87	79.48	–
DeepCNN-BiGRU-Attention	22.89	44.07	61.96	–	36.82	66.23	84.40	–	20.79	39.74	55.97	–	32.64	59.62	78.39	–
STResNet	18.95	36.83	53.52	–	31.45	59.09	79.07	–	21.59	41.40	57.99	–	33.15	60.26	79.01	–
BHS	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A
	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B
	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C

of the models reached Grade C even without weight decay.

B. Responses to Research Questions

a) RQ1: Yes, there is a difference of BP estimation performance between calibration-based and calibration-free subset. The trend was slightly different with and without weight decay. In the case of the calibration-based subset, all models without weight decay showed improved performance compared to conditions with weight decay. On the other hand, in the case of the calibration-free subset, six deep-learning models (m-LRCN, IMSF-Net, ResNet18, ResGRUAttention, DeepCNN-BiGRU-Attention, and ST-ResNet) with weight decay showed improved performance compared to conditions without weight decay. Conversely, three deep-learning models (PP-Net, mPP-Net, and BP-CRNN) were more accurate without weight decay. The reason for this is that more complex models (e.g., ST-ResNet) are more expressive and are overfitting to the training subset in the condition of without weight decay, Thus, in the absence of weight decay, the calibration-based subset, comprising data from the same subjects as the training set, exhibits greater accuracy. In addition, with weight decay, the calibration-free subset, devoid of subjects present in the training data, demonstrates higher accuracy due to the heightened impact of overfitting.

b) RQ2: ST-ResNet showed the highest performance with and without weight decay for the calibration-based subset. For the calibration-free subset, it showed the highest performance with weight decay, while the performance decreased without weight decay. Thus, ST-ResNet is the deep learning model performing the best performance under the calibration-based and calibration-free subset if we can correctly avoid overfitting.

C. Future Work and Limitation

We evaluated the regression model to estimate the SBP and DBP values. Recently, as the other approaches, various methods exist to directly estimate arterial blood pressure (ABP) values from PPG signals [13], [9]. In the future, we

will compare these approaches using a unified dataset.

In this study, the deep-learning models achieved Grades B and C for DBP values in calibration-based subset without weight decay. Given the practical applications, using a model that can be used only a calibration-based subset would require acquiring the user's training data in advance and using it to train the model. Since this is time-consuming for the user, it is desirable to construct a model with a calibration-free subset that does not require prior data acquisition. Therefore, model comparisons should be performed by verifying the accuracy of both calibration-based and calibration-free subset, and by taking measures to prevent overfitting, such as adding a weight decay. Deep-learning models need to be improved to achieve Grade A in this situation.

In addition to differences in subjects, there may be other effects such as day-to-day differences in the way they wear the device when considering its use in a wearable device. To address this challenge, some researchers have proposed methods such as transfer learning [15], continual learning [24], or adding demographic information to the model [19]. In the future, we plan to develop models that incorporate these techniques and evaluate their performance.

IV. CONCLUSIONS

In this study, we compared the performance of nine deep learning models for the estimation of the SBP and DBP values from PPG signals. The results showed that ST-ResNet was the best model for both calibration-based subset and calibration-free subset under the condition that weight decay was included to suppress overfitting. Furthermore, the findings revealed the need to take measures to suppress overfitting for practical applications. In the future, we plan to focus on incorporating techniques such as transfer learning, continual learning, and demographic information to enhance the performance of the model. In summary, our key findings for evaluating the BP estimation model are below.

1. The model should be evaluated with both calibration-based and calibration-free subset.
2. Considering the practical application, the model training should be prevented from the overfitting by using a weight decay, etc.

REFERENCES

[1] F. D. Fuchs and P. K. Whelton, "High blood pressure and cardiovascular disease," *Hypertension*, vol. 75, no. 2, pp. 285–292, 2020.

[2] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, C. A. M. Anderson, P. Arora, C. L. Avery, C. M. Baker-Smith, A. Z. Beaton, A. K. Boehme, A. E. Buxton, *et al.*, "Heart disease and stroke statistics —2023 update: a report from the American Heart Association," *Circulation*, vol. 147, no. 8, pp. e93–e621, 2023.

[3] D. Levy, M. G. Larson, R. S. Vasan, W. B. Kannel, and K. K. L. Ho, "The progression from hypertension to congestive heart failure," *JAMA*, vol. 275, no. 20, pp. 1557–1562, 1996.

[4] E. M. Frese, A. Fick, and S. H. Sadowsky, "Blood pressure measurement guidelines for physical therapists," *Cardiopulmonary Phys. Ther. J.*, vol. 22, no. 2, pp. 5–12, 2011.

[5] X. He, R. A. Goubran, and X. P. Liu, "Evaluation of the correlation between blood pressure and pulse transit time," in *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 17–20, IEEE, 2013.

[6] M. Panwar, A. Gautam, D. Biswas, and A. Acharyya, "PP-Net: A deep learning framework for PPG-based blood pressure and heart rate estimation," *IEEE Sens. J.*, vol. 20, no. 17, pp. 10000–10011, 2020.

[7] C.-T. Yen, J.-X. Liao, and Y.-K. Huang, "Applying a deep learning network in continuous physiological parameter estimation based on photoplethysmography sensor signals," *IEEE Sens. J.*, vol. 22, no. 1, pp. 385–392, 2021.

[8] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network," *Sensors*, vol. 19, no. 15, 2019.

[9] T. Athaya and S. Choi, "An estimation method of continuous non-invasive arterial blood pressure waveform using photoplethysmography: A U-Net architecture-based approach," *Sensors*, vol. 21, no. 5, 2021.

[10] X. F. Teng and Y. T. Zhang, "Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, vol. 4, pp. 3153–3156, IEEE, 2003.

[11] S. C. Gao, P. Wittek, L. Zhao, and W. J. Jiang, "Data-driven estimation of blood pressure using photoplethysmographic signals," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 766–769, IEEE, 2016.

[12] R. He, Z.-P. Huang, L.-Y. Ji, J.-K. Wu, H. Li, and Z.-Q. Zhang, "Beat-to-beat ambulatory blood pressure estimation based on random forest," in *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 194–198, IEEE, 2016.

[13] 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, 2021.

[14] 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, 2023.

[15] J. Leitner, P.-H. Chiang, and S. Dey, "Personalized blood pressure estimation using photoplethysmography: A transfer learning approach," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 1, pp. 218–228, 2021.

[16] C.-T. Yen and C.-H. Liao, "Blood pressure and heart rate measurements using photoplethysmography with modified LRCN.," *Comput. Mater. Contin.*, vol. 71, no. 1, 2022.

[17] D. Wang, Y. Ye, B. Zhang, J. Sun, and C. Zhang, "IMSF-Net: An improved multi-scale information fusion network for PPG-based blood pressure estimation," *Biomed. Signal Process. Control*, vol. 90, 2024.

[18] 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 (CVPR)*, pp. 770–778, 2016.

[19] 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 and Biology Society (EMBC)*, pp. 1–4, IEEE, 2023.

[20] H. Eom, D. Lee, S. Han, Y. S. Hariyani, Y. Lim, I. Sohn, K. Park, and C. Park, "End-to-end deep learning architecture for continuous blood pressure estimation using attention mechanism," *Sensors*, vol. 20, no. 8, 2020.

[21] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11534–11542, 2020.

[22] E. O'Brien, J. Petrie, W. Littler, M. De Swiet, P. L. Padfield, K. O'Malley, M. Jamieson, D. Altman, M. Bland, and N. Atkins, "The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems," *J. Hypertens.*, vol. 8, no. 7, pp. 607–619, 1990.

[23] S. Kanoga, T. Hoshino, S. Kamei, T. Kobayashi, T. Ohmori, M. Uchiyama, and M. Tada, "Comparison of seven shallow and deep regressors in continuous blood pressure and heart rate estimation using single-channel photoplethysmograms under three evaluation cases," *Biomed. Signal Process. Control*, vol. 85, 2023.

[24] C. Zhang, W. Wang, X. Song, Y. Lin, Y. Chen, and X. Ding, "Continual learning for cuffless blood pressure estimation," *Biomed. Signal Process. Control*, vol. 92, 2024.