

Ensemble deep learning for blood pressure estimation using facial videos

Wei Liu¹, Bingjie Wu¹, Menghan Zhou¹, Xingjian Zheng¹, Xingyao Wang¹,
Yiping Xie², Chaoqi Luo³ and Liangli Zhen^{1,*}

¹*Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore*

²*College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.*

³*School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China.*

Abstract

Blood pressure (BP) estimation is a standard and critical component of routine health assessment, especially for cardiac disease patients. Traditional methods typically require direct contact with the patient, which can cause discomfort and inconvenience. Remote photoplethysmography (rPPG) that enables non-contact measurement of the blood volume pulse using trivial cues from facial videos has drawn attention to measure vital signs. This paper presents an ensemble deep learning approach for estimating BP remotely using facial videos. Specifically, to address the vulnerabilities and biases in deep learning models for BP measurement, we emphasize both the accuracy of individual models and the diversity within the ensemble. We utilize advanced deep learning architectures to construct several regression models incorporating convolutional neural networks and transformer blocks, which learn the spatiotemporal relationships between different frames and locations. These trained models are then combined to measure BP readings. Additionally, to enhance the system's robustness under varying lighting conditions, data augmentation techniques are employed to generate more training data. The proposed method is tested on an unseen dataset and the average root of mean squared error (RMSE) is 12.95 mmHg, ranking 1st in the 3rd Vision-based Remote Physiological Signal Sensing (RePSS) Challenge.

Keywords

Blood pressure measurement, remote photoplethysmography, deep learning, ensemble learning

1. Introduction

Blood pressure (BP) measurement is a fundamental diagnostic tool in medical practice, serving as a crucial indicator of cardiovascular health. For instance, elevated BP, or hypertension, is a significant risk factor for cardiovascular diseases, including stroke, heart attack, and renal failure, making accurate and timely measurement vital for early detection and management [1]. The golden standard of continuous BP monitoring is invasive arterial pressure monitoring, which is mainly adopted for primary care [2]. In addition, there are traditional noninvasive BP measurement methods rely on cuffs, but it increases discomfort for patients receiving long-term or frequent monitoring and discouraging real-time measurement [3].

The 3rd Vision-based Remote Physiological Signal Sensing (RePSS) Challenge & Workshop, August 03–09, 2024, Jeju, South Korea

*Corresponding author: Liangli Zhen (email: llzhen@outlook.com)

✉ liuw2@ihpc.a-star.edu.sg (W. Liu); wu_bingjie@ihpc.a-star.edu.sg (B. Wu); zhou_menghan@ihpc.a-star.edu.sg (M. Zhou); zheng_xingjian@ihpc.a-star.edu.sg (X. Zheng); wang_xingyao@ihpc.a-star.edu.sg (X. Wang); yipingx1123@gmail.com (Y. Xie); chaoqiluo7@gmail.com (C. Luo); llzhen@outlook.com (L. Zhen)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Currently, cuffless BP monitor methods have been explored for real-time measurement, and providing convenience and comfort. There are two main methods: pulse transit time (PTT) and pulse wave analysis (PWA) [4]. PTT requires two simultaneous physiological signals to calculate, such as electrocardiography (ECG), phonocardiography (PCG) and seismocardiography (SCG). Compared to PTT, PWA exclusively extracts features from PPG to estimate BP. In recent years, machine learning and deep learning are also employed to establish the mapping relations between PPG and BP [5, 6, 7].

Note that these methods are contact-based and require specific devices like smart watches to make the measurement. Over the past years, remote PPG (rPPG) techniques have been developed for vital sign measurements, especially for heart rate (HR) estimation [8, 9, 10, 11]. Compared to PPG techniques, rPPG-based methods are contactless and can work with digital cameras which are easily accessible nowadays. Other than HR estimation, rPPG techniques have also been applied to make BP estimation from facial videos [12, 13]. While rPPG provides a convenient and cost-effective method for BP estimation, its accuracy can be easily affected by factors such as lighting conditions, skin tones, and motion blur, making rPPG-based BP measurement extremely challenging.

This paper proposes to achieve rPPG-based BP measurement with ensemble deep learning using facial videos. Specifically, to address the vulnerabilities and biases in deep learning models for BP measurement, we prioritize the accuracy of individual models and the diversity within the ensemble. We construct individual regression models by adding a regression head to CNN- and transformer-based backbones. For training each model, we use not only the original RGB images but also features obtained by transforming the color space from RGB to YUV. To enhance the models' performance under varying lighting conditions, data augmentation techniques are employed. Finally, an aggregator is used to combine the outputs from these individual models.

2. Related Work

2.1. Invasive BP Monitoring

Invasive BP estimation can provide continuous and accurate monitoring and is there essential in certain clinical settings, particularly for patients under critical care or during surgery [14, 2]. This method involves the insertion of a catheter into a suitable artery, commonly the radial or femoral artery [15]. The catheter is connected to a pressure transducer, which converts the mechanical pressure exerted by the blood into an electrical signal that can be continuously displayed and monitored. In general, invasive methods provide accurate and continuous monitoring of BP but are only used in certain circumstances due to the significant discomforts to patients.

2.2. Cuff-based BP Estimation

Cuff-based BP measurement is the most common non-invasive method used in both clinical and home settings to assess ABP [16, 3]. This technique utilizes a sphygmomanometer, which includes a cuff that is wrapped around the upper arm and inflated to constrict blood flow. As the cuff deflates, measurements are taken either manually by auscultation—listening to the Korotkoff sounds through a stethoscope—or automatically by oscillometric monitors that

detect blood flow vibrations [17]. Cuff-based methods provide the convenience of quick and easy readings and has been extensively validated for clinical use. However, they impose light discomforts to patients and accuracy can be easily affected by factors like cuff size, arm position, and patient movement.

2.3. PPG-based BP Estimation

PPG-based BP estimation is getting more widely used as the emergence of deep learning algorithms and PPG sensors that can be placed on the finger, earlobe, or over the wrist [18]. Variations in light absorption during the cardiac cycle are measured, providing information about the blood flow, heart rate, and other cardiovascular attributes. By analyzing these variations, algorithms can estimate systolic and diastolic BP values [6, 19]. PPG-based methods offer ease of use, the potential for continuous monitoring, and the absence of discomfort of cuff methods. However, the accuracy is sensitive to motion artifacts and changes in sensor placement.

2.4. The rPPG-based BP Estimation

Recently, the rPPG-based methods offer a non-contact way for BP estimation by using video cameras to detect blood volume changes in facial skin [20]. This technology, which can be implemented with standard RGB cameras found in common devices like smartphones and tablets, captures subtle changes in light reflection off the skin due to pulsating blood flows [21, 22, 23]. The rPPG-based methods are non-invasive and use widely-accessible cameras, making it potentially cost-effective and convenient for regular BP checks. However, the accuracy can be compromised by factors such as motion and variable lighting conditions, posing challenges for its use in dynamic or uncontrolled environments.

3. Methodology

The overall framework of our ensemble deep learning method is illustrated in Fig. 1, from which we can see that there are multiple regression models. To import diversity, multiple models are trained using different input feature vectors, backbones, or random seeds. The outputs of individual models are then fused with an aggregator.

3.1. Data Preprocessing

A short clip is extracted from the original full video and then partitioned into frames. It is worth pointing out that we select the clip closest to the time when BP is measured to mitigate the impact of BP fluctuation during video taking. If the video is recorded before BP measurement, the last part the video is selected and vice versa for videos taken after BP measurement. The face region of each frame is then cropped and resized to 128×128 . To improve model performance in different lighting conditions, data augmentation technique is applied during the training process.

It has been demonstrated in [11, 24] that alternative color spaces derived from RGB videos are beneficial for better representation of HR signal. Other than original RGB images, we also

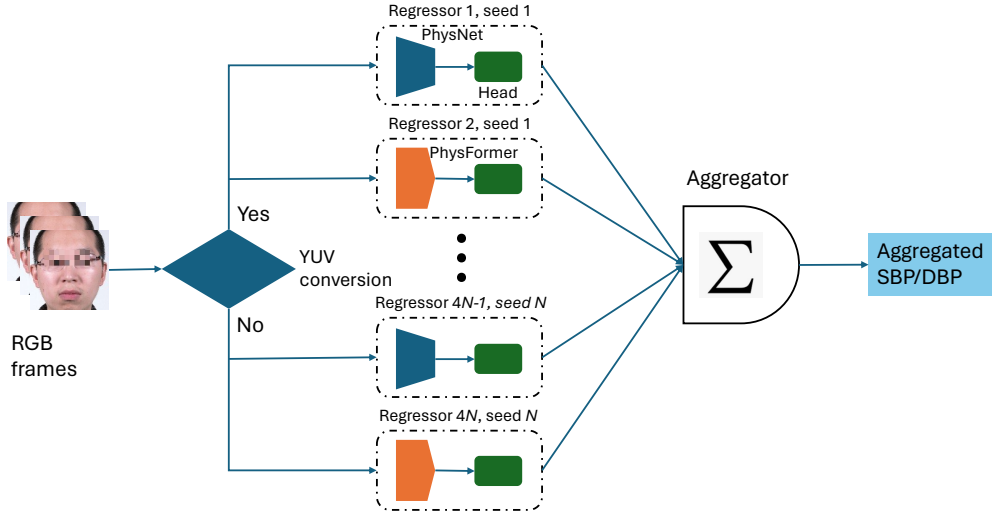


Figure 1: Overall framework of the proposed method. SBP: systolic BP. DBP: diastolic BP.

explored using YUV color space for BP estimation. Mathematically, the transformation from RGB to YUV can be calculated as

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

where R , G , and B represent the red, green, and blue color components of an image, respectively. Y represents the luminance component, while U and V represent the chrominance components, capturing the color information minus the brightness.

3.2. Network Structure

3.2.1. Backbones

We utilize two state-of-the-art models as the backbone for our BP estimation model, including a 3D CNN model named PhysNet [25] and a transformer-based model named PhysFormer [8]. The output of two backbones are both the estimated PPG signal which has been used to recover ABP [6, 26, 27]. Therefore, we keep all the layers of the backbones so that the output of the backbone remain as the PPG signal. The output of the backbone is a 1D signal that has the same length as the number of input frames. The details of the backbones can be found in [25, 8].

3.2.2. Regression head

We stack a regression head with one hidden layer on top of the backbone and the regression head has two output nodes corresponding to SBP and DBP, respectively. The regression head

can be formulated as

$$\begin{aligned}\mathbf{h} &= \sigma(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) \\ \mathbf{y} &= \mathbf{W}^{(2)}\mathbf{h} + \mathbf{b}^{(2)}\end{aligned}\quad (2)$$

where σ is the standard Sigmoid function. \mathbf{W} and \mathbf{b} are the weights and biases, respectively. \mathbf{x} is the output signal from the backbone. \mathbf{h} denotes the vector at the hidden layer. \mathbf{y} denotes the output vector consisting of DBP and SBP.

3.3. Loss Function

The average RMSE of SBP and DBP is used as the loss function to train our models, defined as

$$L = 0.5 \times \sqrt{\frac{\sum_{i=1}^N (g_i^d - y_i^d)^2}{N}} + 0.5 \times \sqrt{\frac{\sum_{i=1}^N (g_i^s - y_i^s)^2}{N}} \quad (3)$$

where g_i^d and g_i^s are the ground-truth DBP and SBP results of the i^{th} sample, respectively. y_i^d and y_i^s are the predicted DBP and SBP of the i^{th} sample, respectively. N is the number of samples.

3.4. Aggregation

As mentioned above, multiple individual models are trained with different input features (RGB or YUV), backbones (PhysNet or PhysFormer) and random seeds to introduce diversity to our ensemble method. Ensemble learning is used to aggregate the outputs of individual models. For each sample, we remove the top- n and bottom- n results and then calculate the average of the rest outputs as

$$\begin{aligned}y_{\text{ens}}^d &= \frac{1}{N - 2n} \sum_{i=n+1}^{N-n} y_i^d \\ y_{\text{ens}}^s &= \frac{1}{N - 2n} \sum_{i=n+1}^{N-n} y_i^s\end{aligned}\quad (4)$$

where y_{ens}^d and y_{ens}^s are the aggregated prediction of DBP and SBP, respectively. y_i^d and y_i^s represent the predicted DBP and SBP of the i^{th} model when they are arranged in ascending order. N is the number of individual models. The top- n and bottom- n values are neglected.

4. Experimental Study

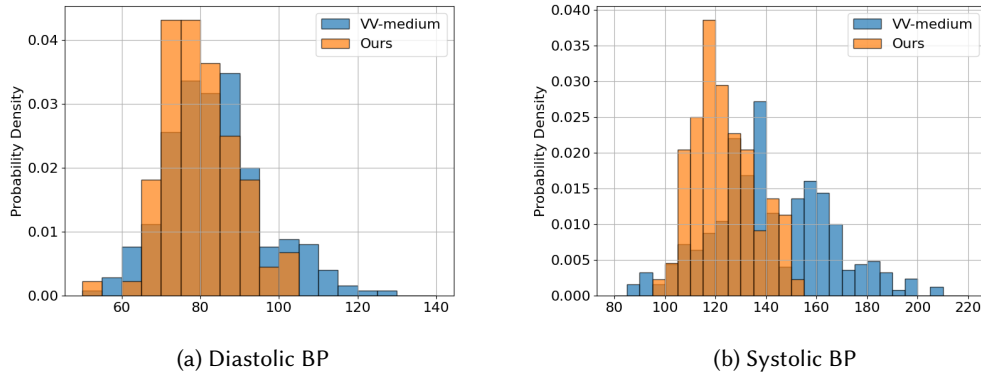
4.1. Experimental Setup

The proposed method is tested using PyTorch on a server equipped with Intel(R) Xeon(R) Gold 6430 CPU and RTX4090 GPU. The models are trained for 150 epochs using AdamW optimizer [28] with learning rate $lr = 1 \times 10^{-5}$ and $weight_decay = 1 \times 10^{-5}$. The value of n in Equ. (4) is set as 3.

Table 1

Brief Summary of Datasets for Training. FPS: Frames Per Second.

Dataset	# Subject	# Videos	# BP labels	Video length (s)	FPS
VV-medium [29]	250	499	250	30	30
Our private dataset	88	88	88	120	30

**Figure 2:** Distribution of diastolic and systolic BP of two datasets used for training.

4.2. Datasets

Two datasets are used for model training and validation, including the VV-medium dataset [29] and our private dataset. A brief summary of these two datasets is reported in Table 1 and distribution are illustrated in Fig. 2. VV-medium dataset [29] has more videos than BP label because each BP label corresponds to multiple videos. It is shown that BP of VV-medium dataset [29] is more diversely distributed compared to our dataset.

For testing, the OBF Database – Oulu BioFace Database [30, 31] consisting of 100 subjects and 200 facial videos with DBP/SBP labels is used for evaluation. Note that for testing, we only have access to the facial videos and have no access to the ground-truth BP labels.

4.3. Evaluation Metrics

Three metrics, including the root of mean squared error (RMSE), mean absolute error (MAE) and Pearson correlation coefficient r , are used to evaluate model performance on the validation dataset, which are defined as

$$\begin{aligned}
\text{RMSE} &= \sqrt{\frac{\sum_{i=1}^N (g_i - y_i)^2}{N}} \\
\text{MAE} &= \frac{1}{N} \sum_{i=1}^N |g_i - y_i| \\
r &= \frac{\sum_{i=1}^N (g_i - \bar{g})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (g_i - \bar{g})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}
\end{aligned} \tag{5}$$

where g and y are the ground-truth and predicted SBP/DBP, respectively. N is the number of samples. \bar{g} and \bar{y} indicate the average values of ground-truth and prediction, respectively.

For testing, the average RMSE of DBP and SBP is used to evaluate model performance, calculated as

$$\text{RMSE}_{\text{avg}} = 0.5 \times \text{RMSE}_d + 0.5 \times \text{RMSE}_s \tag{6}$$

where RMSE_d and RMSE_s are the RMSE of DBP and SBP, respectively.

4.4. Experimental Results

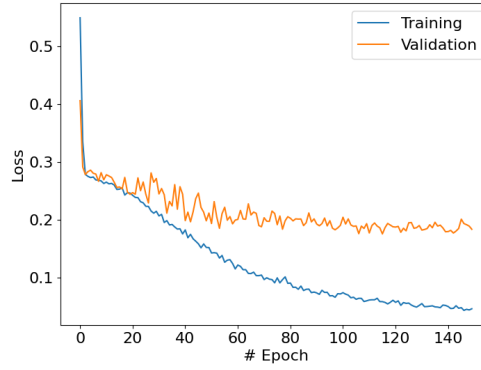


Figure 3: The training and validation loss over epochs during training.

We randomly split the available dataset into training dataset (80%) and validation dataset (20%). The learning curve one of our individual models is shown in Fig. 3 and it shows that the model converges well in 150 epochs. The scatter plots on validation set are illustrated in Fig. 4 where RMSE, MAE and r are also reported. It is shown that the estimated and true BP are strongly correlated and the errors of most samples are within ± 10 mmHg.

The RMSE of DBP and SBP are 8.93 mmHg and 11.03 mmHg, respectively. The MAE and RMSE of SBP are larger than those of DBP because SBP range is larger and more diversely distributed, which can be seen from Fig. 2. On the testing dataset, the average RMSE is 12.95 mmHg and a comparison is reported in Table 2. One can see that our method outperforms competing methods by more than 0.5 mmHg.

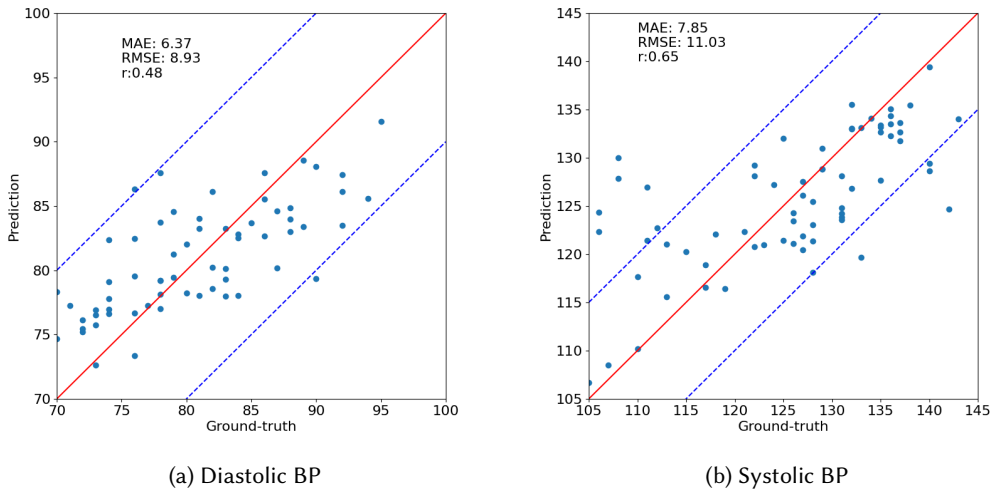


Figure 4: Scatter plots of ground-truth and predicted BP. The solid red line indicates that prediction is the same as the ground-truth. The dashed blue lines indicate error of ± 10 mmHg.

Table 2

A Comparison of Our Method and the Peer Methods [32].

Rank	Team	RMSE _{avg}
1	Face AI (BP)-Ours	12.95
2	PCA_Vital	13.48
3	Rythm	13.59
4	SCUT_rPPG	15.06
5	IAI-USTC	16.01
6	NeuroAI	16.56

5. Conclusion

This paper presented an ensemble deep learning method for BP estimation using facial videos. To improve the diversity of models to ensemble learning, multiple models are built with different backbones and input feature vectors. Besides, data augmentation technique is used to improve model performance under different lighting conditions. The outputs of individual models are fused with an aggregator. Our method is tested on an unseen dataset in the RePSS challenge, and the average RMSE of SBP and DBP is 12.95 mmHg, which outperforms all the peer methods and indicates the effectiveness of our proposed method.

References

- [1] F. D. Fuchs, P. K. Whelton, High blood pressure and cardiovascular disease, *Hypertension* 75 (2020) 285–292.

- [2] B. Saugel, K. Kouz, A. S. Meidert, L. Schulte-Uentrop, S. Romagnoli, How to measure blood pressure using an arterial catheter: a systematic 5-step approach, *Critical Care* 24 (2020) 1–10.
- [3] D. S. Picone, M. G. Schultz, P. Otahal, S. Aakhus, A. M. Al-Jumaily, J. A. Black, W. J. Bos, J. B. Chambers, C.-H. Chen, H.-M. Cheng, et al., Accuracy of cuff-measured blood pressure: systematic reviews and meta-analyses, *Journal of the American College of Cardiology* 70 (2017) 572–586.
- [4] R. Mukkamala, J.-O. Hahn, A. Chandrasekhar, Photoplethysmography in noninvasive blood pressure monitoring, in: *Photoplethysmography*, Elsevier, 2022, pp. 359–400.
- [5] D. Konstantinidis, P. Iliakis, F. Tatakis, K. Thomopoulos, K. Dimitriadis, D. Tousoulis, K. Tsioufis, Wearable blood pressure measurement devices and new approaches in hypertension management: the digital era, *Journal of human hypertension* 36 (2022) 945–951.
- [6] N. Ibtehaz, S. Mahmud, M. E. Chowdhury, A. Khandakar, M. Salman Khan, M. A. Ayari, A. M. Tahir, M. S. Rahman, Ppg2abp: Translating photoplethysmogram (ppg) signals to arterial blood pressure (abp) waveforms, *Bioengineering* 9 (2022) 692.
- [7] K. R. Vardhan, S. Vedanth, G. Poojah, K. Abhishek, M. N. Kumar, V. Vijayaraghavan, Bp-net: Efficient deep learning for continuous arterial blood pressure estimation using photoplethysmogram, in: *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2021, pp. 1495–1500.
- [8] Z. Yu, Y. Shen, J. Shi, H. Zhao, P. H. Torr, G. Zhao, Physformer: Facial video-based physiological measurement with temporal difference transformer, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 4186–4196.
- [9] Z. Yu, Y. Shen, J. Shi, H. Zhao, Y. Cui, J. Zhang, P. Torr, G. Zhao, Physformer++: Facial video-based physiological measurement with slowfast temporal difference transformer, *International Journal of Computer Vision* 131 (2023) 1307–1330.
- [10] X. Liu, B. Hill, Z. Jiang, S. Patel, D. McDuff, Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement, in: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 5008–5017.
- [11] H. Shao, L. Luo, J. Qian, S. Chen, C. Hu, J. Yang, Tranphys: Spatiotemporal masked transformer steered remote photoplethysmography estimation, *IEEE Transactions on Circuits and Systems for Video Technology* (2023).
- [12] H. Luo, D. Yang, A. Barszczyk, N. Vempala, J. Wei, S. J. Wu, P. P. Zheng, G. Fu, K. Lee, Z.-P. Feng, Smartphone-based blood pressure measurement using transdermal optical imaging technology, *Circulation: Cardiovascular Imaging* 12 (2019) e008857.
- [13] Y. Zhou, H. Ni, Q. Zhang, Q. Wu, The noninvasive blood pressure measurement based on facial images processing, *IEEE Sensors Journal* 19 (2019) 10624–10634.
- [14] H. L. Li-wei, M. Saeed, D. Talmor, R. Mark, A. Malhotra, Methods of blood pressure measurement in the icu, *Critical care medicine* 41 (2013) 34–40.
- [15] S. Romagnoli, Z. Ricci, D. Quattrone, L. Tofani, O. Tujjar, G. Villa, S. M. Romano, A. R. De Gaudio, Accuracy of invasive arterial pressure monitoring in cardiovascular patients: an observational study, *Critical care* 18 (2014) 1–11.
- [16] P. Palatini, R. Asmar, Cuff challenges in blood pressure measurement, *The Journal of Clinical Hypertension* 20 (2018) 1100–1103.
- [17] M. Forouzanfar, H. R. Dajani, V. Z. Groza, M. Bolic, S. Rajan, I. Batkin, Oscillometric blood

- pressure estimation: past, present, and future, *IEEE reviews in biomedical engineering* 8 (2015) 44–63.
- [18] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, H. Nazeran, A review on wearable photoplethysmography sensors and their potential future applications in health care, *International journal of biosensors & bioelectronics* 4 (2018) 195.
 - [19] M. Panwar, A. Gautam, D. Biswas, A. Acharyya, Pp-net: A deep learning framework for ppg-based blood pressure and heart rate estimation, *IEEE Sensors Journal* 20 (2020) 10000–10011.
 - [20] Y. Lu, C. Wang, M. Q.-H. Meng, Video-based contactless blood pressure estimation: A review, in: *2020 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, IEEE, 2020, pp. 62–67.
 - [21] F. Schruppf, P. Frenzel, C. Aust, G. Osterhoff, M. Fuchs, Assessment of deep learning based blood pressure prediction from ppg and rppg signals, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3820–3830.
 - [22] B.-F. Wu, B.-J. Wu, B.-R. Tsai, C.-P. Hsu, A facial-image-based blood pressure measurement system without calibration, *IEEE Transactions on Instrumentation and Measurement* 71 (2022) 1–13.
 - [23] Y. Chen, J. Zhuang, B. Li, Y. Zhang, X. Zheng, Remote blood pressure estimation via the spatiotemporal mapping of facial videos, *Sensors* 23 (2023) 2963.
 - [24] X. Niu, S. Shan, H. Han, X. Chen, Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation, *IEEE Transactions on Image Processing* 29 (2019) 2409–2423.
 - [25] Z. Yu, X. Li, G. Zhao, Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks, in: *Proceedings of the British Machine Vision Conference*, 2019.
 - [26] M. A. Mehrabadi, S. A. H. Aqajari, A. H. A. Zargari, N. Dutt, A. M. Rahmani, Novel blood pressure waveform reconstruction from photoplethysmography using cycle generative adversarial networks, in: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2022, pp. 1906–1909.
 - [27] L. N. Harfiya, C.-C. Chang, Y.-H. Li, Continuous blood pressure estimation using exclusively photoplethysmography by lstm-based signal-to-signal translation, *Sensors* 21 (2021) 2952.
 - [28] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, *arXiv preprint arXiv:1711.05101* (2017).
 - [29] P.-J. Toye, Vital videos: A dataset of videos with ppg and blood pressure ground truths, *arXiv preprint arXiv:2306.11891* (2023).
 - [30] X. Li, I. Alikhani, J. Shi, T. Seppanen, J. Junntila, K. Majamaa-Voltti, M. Tulppo, G. Zhao, The obf database: A large face video database for remote physiological signal measurement and atrial fibrillation detection, in: *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, IEEE, 2018, pp. 242–249.
 - [31] Z. Yu, W. Peng, X. Li, X. Hong, G. Zhao, Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement, in: *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 151–160.
 - [32] Z. Sun, The 3rd repss track 2, 2024. URL: <https://kaggle.com/competitions/the-3rd-repss-t2>.