# Video-based remote blood pressure measurement using convolutional networks and random forest

Wei Zhuo, Jianjun Qian\*, Hang Shao, Lei Luo and Jian Yang

PCA Lab, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China

#### Abstract

Blood pressure (BP) is an important vital sign that is highly correlated with human health. With the development and maturity of remote photoplethysmograpy (rPPG) technology, the analysis of facial video makes it possible to measure BP in a non-contact way. In this paper, we propose a network for remote BP measurement, named RBP-CNN. Specifically, we first extract blood volume pulse (BVP), heart rate (HR), age and body mass index (BMI) from the facial video and analyze their correlation with BP during which we find a close correlation between diastolic blood pressure (DBP) and systolic blood pressure (SBP). Then, RBP-CNN is designed based on residual convolution, local and global attention mechanism, to extract the implicit BP-related features, which are hard to be discovered and manually extracted. Finally, we use the ensemble learning algorithm random forest (RF) to fuse these features to measure BP and verify our method by RF's feature importance. Our approach is trained and tested on 322 and 200 samples provided by Track 2 of the challenge respectively, and it achieves the root mean squared error (RMSE) of 13.48281 which ranks second in the final leaderboard. The codes are publicly available at https://github.com/zhuowei123/3rd-RePSS-track2.git

#### **Keywords**

RePSS, remote photoplethysmograpy, ensemble learning, blood pressure estimation

### 1. Introduction

Blood pressure (BP) is an important vital sign in diagnosing certain cardiovascular diseases such as hypertension [1, 2, 3]. There are two kinds of BP in the human body, namely diastolic blood pressure (DBP) and systolic blood pressure (SBP), which represent the pressure of blood on blood vessels during contraction and relaxation of the heart. In real life, BP is usually measured by contact detection instruments or wearable medical devices. Auscultation is the most traditional method of BP measurement which can well determine the BP state at the time of measurement but it's often influenced by the experience of the auscultator and the environment, resulting in measurement errors. Although cuff oscillography can overcome some shortcomings of auscultation, the inflatable cuff tends to bring uncomfortable experience to the personnel being tested. Therefore, it is of great significance to study convenient and accurate



IJCAI 2024: International Joint Conference on Artificial Intelligence, August 3–9, 2024, Jeju, South Korea \*Corresponding author.

<sup>🛆</sup> weizhuo@njust.edu.cn (W. Zhuo); csjqian@njust.edu.cn (J. Qian); shaohang@njust.edu.cn (H. Shao); cslluo@njust.edu.cn (L. Luo); csjyang@njust.edu.cn (J. Yang)

https://github.com/zhuowei123/ (W. Zhuo)

O009-0007-3109-1290 (W. Zhuo); 0000-0002-0968-8556 (J. Qian); 0000-0002-2452-6985 (H. Shao);
 0000-0002-9976-0442 (L. Luo); 0000-0003-4800-832X (J. Yang)

<sup>© 2024</sup> Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

non-contact BP as well as other physiological signals measurement for health monitoring. In order to solve the discomfort and inconvenience caused by contact measuring equipment. Remote photoplethysmography (rPPG) [4, 5, 6, 7, 8] methods are developing fast in recent years, which aim to measure heart activity remotely without any contact, makes non-contact physiological signal measurement possible. To study more robust computer vision algorithms and biomedical signal processing methods for extracting physiological signals from facial videos, the 3rd Vision-based Remote Physiological Signal Sensing (RePSS) workshop will be held in conjunction with the International Joint Conference on Artificial Intelligence (IJCAI 2024). There are two tracks in The 3rd RePSS challenge, and the task of the Track2 is facial video-based BP measurement.

BP is closely related to various physiological signals, among which the pulse transit time (PTT) [9] is the most representative. To be specific, PTT refers to the time for a pulse to travel between two different body parts. According to whether PTT is used, contactless BP measurement methods can be divided into PTT-based methods and None-PTT methods. Both types of methods have their limitations, PTT-based methods have high requirements for video frame rate, content and stability. The None-PTT methods are vulnerable to individual differences.

Considering the feasibility of BP measurement based on facial videos, our approach focus on the analysis of physiological signals in the facial video. To be specific, blood volume pulse (BVP), heart rate (HR), body mass index (BMI) and age are extracted from the facial video and frame. Then, RBP-CNN captures the high-dimensional features from BVP to measure the dynamic information in the facial video. Finally, the ensemble algorithm random forest (RF) is used for feature fusion and BP measuring.

### 2. Related works

**PTT-based methods** Remote BP measurement is sensitive to head shaking [10, 11]. Noncontact PTT based BP measurements often require video of multiple parts or other signal support to improve robustness. For example, Fan *et al.* [12] extract a palm-to-face PTT from the video and feeds it into a physical BP model. Wu *et al.* [13] employ PTT from two face regions and fuses heart rate variability (HRV), BMI and BVP into a multi-modal model for BP measurement.

**None-PTT methods** Different from the PTT-based methods, the None-PTT methods measure BP by fusing physiological signals such as BVP, HR, HRV, BMI, and age. Zhou *et al.* [14] input the peaks and troughs of the BVP waveform into a linear regression model to predict BP. Rong *et al.* [15] extract 26 features from BVP for BP estimation, and train them through four machine learning algorithms. In addition to BVP features, Luo *et al.* [16] take 29 meta-features(room temperature, subjects'ages, weight, etc.) into account to estimate BP.

### 3. Methodology

Our method can be divided into two stages: RBP-CNN training to extract the BP-related feature from BVP and RF training for multi-feature fusion and BP measuring. In this section we will detail each of these stages in turn.

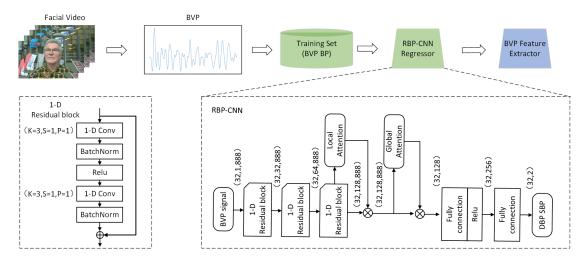


Figure 1: the process of BVP features extraction using RBP-CNN.

#### 3.1. RBP-CNN for BVP feature extraction

As shown in Figure 1, the first stage of our method is BVP feature extraction using RBP-CNN. In this part, we'll introduce in the following order: BVP extraction, principle and structure of RBP-CNN, and loss function.

Nowadays, there are already many excellent unsupervised [17, 18, 19, 20, 21] and supervised [22, 23, 24, 25] methods that can represent BVP signals from facial videos. Robust pulse rate from chrominance-based rppg (CHROM) [18] is a traditional and effective unsupervised method, which is used in our BVP extraction. The extracted BVP signal is a one-dimensional time series that changes with time. In previous studies, BVP signal are often used to measure physiological signals such as HR and HRV. However, we believe that in addition to these important reference indicators in medicine, there are high-dimensional features related to BP implied in BVP.

We design RBP-CNN based on residual convolution, local and global attention mechanism to learn features of BVP signals. ResNet [26] has a strong feature representation ability with residual connections and it is widely used in time series analysis. The local attention mechanism is used to focus on local regions within sequence data dynamically and selectively. In the context of time series data, the local attention mechanism enables the model to adjust its attention based on specific parts of the input sequence, allowing for more effective capture of key information within the sequence. The global attention mechanism is employed to consider information from the entire input sequence when making predictions or feature representation. When processing time series data, the global attention mechanism enables the model to weight all time steps equally, allowing it to capture global patterns and relationships within the sequence.

As illustrated in Figure 1, RBP-CNN consists of three 1D residual blocks (depicted in the left part of Figure 1), local attention, global attention and two fully connected layers. We feed BVP signal and BP into RBP-CNN, and BVP is first mapped to high-dimensional feature space through three residual blocks. Then, the weight of each time step of BVP is adjusted and weighed by the local attention and global attention mechanism. Finally, the dimension is reduced by two fully

connected layers, and the BP is predicted.

It's worth noting that BP measurement based on multiple physiological signals is essentially an imbalanced regression [27, 28, 29] task. Because it can be easily observed that most training samples of BP regression concentrate on adults and middle-aged people, while the samples of children and elderly people are fewer, so the labels are imbalanced. To cope with this problem, balanced mean squared error (BMSE) [30] is proposed, which addresses the label imbalance from a statistical perspective, below we give a brief introduction to its principle.

The  $y_{\text{pred}}$  regression can be modeled as a Gaussian distribution, and the mean squared error(MSE) is equivalent to the negative log-likelihood loss of this distribution  $p(y \mid x; \theta)$ . So, training the MSE regression model is equivalent to modeling the distribution.

$$p(\boldsymbol{y} \mid \boldsymbol{x}; \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{y}; \boldsymbol{y}_{\text{pred}}, \sigma_{\text{noise}}^2 \mathbf{I}\right),$$
(1)

where  $\boldsymbol{y}$  is the label,  $\boldsymbol{x}$  is the input,  $\boldsymbol{\theta}$  is the regressor's parameter,  $\boldsymbol{y}_{\text{pred}}$  is the regressor's prediction and  $\sigma_{\text{noise}}$  is the scale of an i.i.d error term  $\mathcal{N}\left(0, \sigma_{\text{noise}}^2 \mathbf{I}\right)$ . In the imbalanced regression task, we train on an imbalanced distribution  $p_{\text{train}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right)$  and test on a balanced distribution  $p_{\text{bal}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right)$ , which leads to a distribution mismatch. By Bayes' rule we get:

$$\frac{p_{\text{train}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right)}{p_{\text{bal}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right)} \propto \frac{p_{\text{train}}\left(\boldsymbol{y}\right)}{p_{\text{bal}}\left(\boldsymbol{y}\right)}$$
(2)

Equation 2 shows that the ratio of  $p_{\text{train}}(\boldsymbol{y} \mid \boldsymbol{x})$  and  $p_{\text{bal}}(\boldsymbol{y} \mid \boldsymbol{x})$  is proportional to  $p_{\text{train}}(\boldsymbol{y})$ , so for less distributed labels, that is, for lower  $p_{\text{train}}(\boldsymbol{y})$ , regressor using MSE will underestimate on rare labels. BMSE assumes that  $p_{\text{train}}(\boldsymbol{y} \mid \boldsymbol{x})$  and  $p_{\text{bal}}(\boldsymbol{y} \mid \boldsymbol{x})$  have same label conditional distribution. Then  $p_{\text{train}}(\boldsymbol{y} \mid \boldsymbol{x})$  can always be expressed by  $p_{\text{bal}}(\boldsymbol{y} \mid \boldsymbol{x})$  and  $p_{\text{train}}(\boldsymbol{y})$  as:

$$p_{\text{train}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right) = \frac{p_{\text{bal}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right) \cdot p_{\text{train}}\left(\boldsymbol{y}\right)}{\int_{Y} p_{\text{bal}}\left(\boldsymbol{y}' \mid \boldsymbol{x}\right) \cdot p_{\text{train}}\left(\boldsymbol{y}'\right) d\boldsymbol{y}'}.$$
(3)

Finally, for a regressor's prediction  $p_{\text{train}}$ , and a training label distribution prior  $p_{\text{train}}(\boldsymbol{y} \mid \boldsymbol{x})$ , the BMSE loss is defined as:

$$L = -\log p_{\text{train}} \left( \boldsymbol{y} \mid \boldsymbol{x}; \boldsymbol{\theta} \right)$$
  
=  $-\log \frac{p_{\text{bal}} \left( \boldsymbol{y} \mid \boldsymbol{x}; \boldsymbol{\theta} \right) \cdot p_{\text{train}} \left( \boldsymbol{y} \right)}{\int_{Y} p_{\text{bal}} \left( \boldsymbol{y}' \mid \boldsymbol{x}; \boldsymbol{\theta} \right) \cdot p_{\text{train}} \left( \boldsymbol{y}' \right) d\boldsymbol{y}'}$   
$$\cong -\log \mathcal{N} \left( \boldsymbol{y}; \boldsymbol{y}_{\text{pred}}, \sigma_{\text{noise}}^2 \mathbf{I} \right)$$
  
 $+\log \int_{Y} \mathcal{N} \left( \boldsymbol{y}'; \boldsymbol{y}_{\text{pred}}, \sigma_{\text{noise}}^2 \mathbf{I} \right) \cdot p_{\text{train}} \left( \boldsymbol{y}' \right) d\boldsymbol{y}',$  (4)

where  $\cong$  hides a constant term  $-\log p_{\text{train}}(\boldsymbol{y})$ . It can be noted that the calculation of BMSE loss involves the calculation of a double integral. For simplicity, we use its batch-based Monte-Carlo (BMC) approximate implementation for the loss calculation of RBP-CNN.

#### 3.2. Multi-feature fusion with Random Forest

As demonstrated in Figure 2, the second stage of our method is multi-feature fusion and BP prediction based on RF. In this part, we will introduce our scheme in the order of feature extraction, feature correlation analysis, and feature fusion.

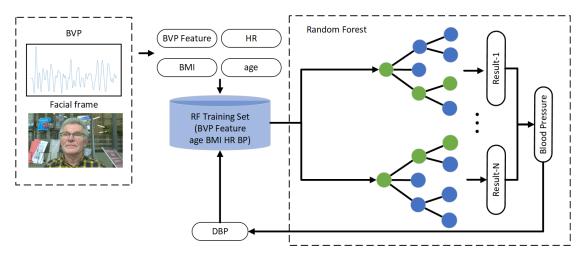


Figure 2: the process of multi-feature fusion and BP measuring using RF.

In medicine, the primary cause of hypertension is arteriosclerosis, and the most directly associated factors with arteriosclerosis are age and BMI, that's why hypertension is more prevalent in obese and middle-aged to elderly populations. Additionally, these individuals are also prone to show abnormalities in HR. Therefore, we take care of age, BMI, and HR as important features for BP measurement. Specifically, we input any frame from the facial video into pre-trained models [31] to estimate age and BMI. Heart rate is calculated from BVP signal by fourier transform.

It can be seen from Figure 3 that the DBP's pearson correlation coefficients with age, BMI, HR are 0.301, 0.238 and 0.133 respectively (p<0.001) among which DBP is moderately correlated with age and weakly correlated with BMI and HR. As shown in Figure 4, the pearson correlation coefficients of SBP with age, BMI and DBP are 0.562, 0.286 (p<0.001) and 0.704 (p<0.05), indicating that they have moderate, weak and strong correlations with SBP respectively. Commonly used physiological information such as age, BMI and HR have been used in the previous works. Researchers have focused on the relationship between them and BP, however, few people pay attention to the internal correlation between DBP and SBP. We notice it and utilise DBP in SBP prediction.

Multi-feature fusion is realized by RF [32], which is a widely used ensemble algorithm. RF is composed of multiple decision trees, and the final prediction result is determined by the voting results of each decision tree. In regression task, the output of each decision tree is a continuous value, and the average of the output results of all decision trees is taken as the final result. RF can deal with high-dimensional and imbalanced datasets, and has the advantages of high accuracy and robustness. At the same time, we can also evaluate the importance of features, which is helpful for us to verify the effectiveness of features through experiments.

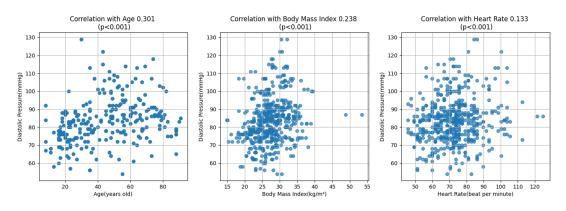


Figure 3: Scatter plots showing the DBP's linear relationship with age, BMI and HR.

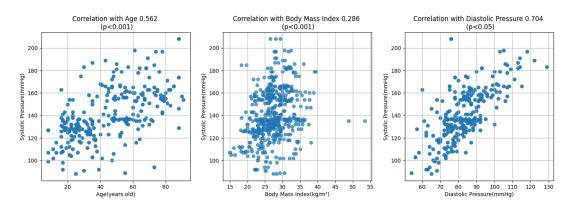


Figure 4: Scatter plots showing the SBP's linear relationship with age, BMI and SBP.

### 4. Experiments

In the 3rd RePSS challenge Track2, we use 322 samples from Vital Videos for training, of which 162 samples are used for training RBP-CNN and 160 for training the random forest. 200 label unknowned samples from OBF are used for testing.

### 4.1. Datasets

Vital Videos (VV) [33] is a public dataset of videos with PPG and BP ground truths, which in total contains information about 900 different participants. For each participant, 2 or 3 30s uncompressed video are collected, along with personal information (gender, age, skin color), PPG, HR, blood oxygen saturation and BP. The dataset includes roughly equal numbers of males and females, as well as participants of all ages, skin color in different locations, ensuring a variety of different background and lighting conditions.

OBF [34] is a large face video database for remote physiological signal measurement and atrial fibrillation (AF) detection. It contains data from 100 healthy individuals as well as six AF

patients. For each participant, multi-modal data (RGB videos, NIR video, ECG, BVP, RF) are recorded simultaneously during two phases, each lasting 5 minutes. For healthy participants and patients with AF, the first and second phases are resting state, post-exercise HR increase, before and after cardioversion treatment respectively.

#### 4.2. Training Procedure

For the RBP-CNN training stage, the BVP is extracted from the corresponding facial video, after which the HR is calculated. Subsequently, the BVP signals and BP labels are fed into RBP-CNN for training. After that, the last fully connected layer of RBP-CNN is removed and it becomes a BVP feature extractor. We then feed BVP into the feature extractor to capture BP-related features. At last, both BVP feature and other physiological information are used to build the RF training set, and then fit the DBP and SBP through the RF regression model.

For the RF regressor training stage, two regressors (DBP regressor and SBP regressor) are trained. The BVP and HR are calculated from the facial video. At the same time, the first frame of each sample's facial video is used for BMI estimation (age is available in VV). Lastly, the features learned from BVP, age, BMI, HR, DBP (SBP regressor training only) are fused to train DBP and SBP RF regressors.

There are 507 samples from 250 participants for training initially. To improve performance across datasets, we use the evaluated age to approximate the distribution of the training set to the test set distribution. Finally, 322 samples from 160 participants are selected, among which 162 and 160 samples are used for training RBP-CNN and RF regressor.

The RBP-CNN model is implemented based on pytorch framework and trained on a NIVIDA GeForce GTX 1650 GPU for 200 epochs with a learning rate of 0.001. The RF regressor is implemented based on sklearn and trained on Intel (R) Core (TM) i5-9300H CPU and the n\_estimators, max\_depth, criterion are set to 1000, 6, absolute\_error respectively.

#### 4.3. Evaluation Metric

In the training process of RBP-CNN model, we use mean absolute error (MAE) and MSE as the model evaluation metric. For the actual test of Track 2, The root mean squared error (RMSE) of the ground truth DBP, SBP with the submitted ones are calculated successively, and then they are averaged as the final race score.

$$RMSE_2 = 0.5\sqrt{\frac{\sum_{i=1}^{N} (s_i - s'_i)^2}{N}} + 0.5\sqrt{\frac{\sum_{i=1}^{N} (d_i - d'_i)^2}{N}}$$
(5)

where  $s_i$  is the ground truth SBP of the ith test sample,  $s'_i$  is the submitted SBP of the ith test sample. Similarly,  $d_i$  and  $d'_i$  are the ground truth DBP and submitted one of the ith test sample.

#### 4.4. Results

As shown in Table 1, our team (PCA\_Vital) achieves second place on the 3nd RePSS Challenge Track 2. The final score of our submitted BP prediction is 13.48281, which is behind first place

Table 1The final leaderboard of The 3rd RePSS Track 2.

Ranking	Team Name	Captain Affiliation	RMSE <sub>BP</sub> (mmHg)
1	Face AI(BP)	Institute of High Performance Computing	12.95258
		Agency for Science, Technology and Research	
2	PCA_Vital (Ours)	Nanjing University of Science and Technology	13.48281
3	Rhythm	University of Science and Technology Beijing	13.59307
4	SCUT_rPPG	South China University of Technology	15.06056
5	IAI-USTC	University of Science and Technology of China	16.01179
6	NeuroAl	Kwangwoon University	16.56091

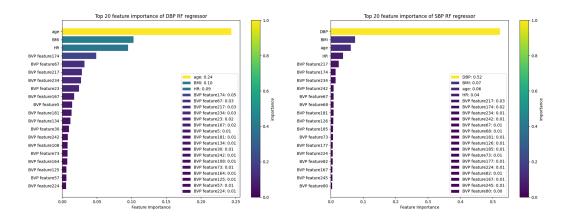


Figure 5: top 20 feature importance of DBP and SBP RF regressors.

0.53023 mmHg and ahead of the third place 0.11026 mmHg but significantly ahead of the fourth place.

Figure 5 shows the top 20 feature importance of DBP and SBP RF regressors. It can be observed that for both DBP and SBP regressors, BMI and age rank among the top three feature importance. Notably, the feature importance of DBP in SBP RF regressor comes to 0.52, which is obviously ahead of other features. It confirms the strong correlation between DBP and SBP. At the same time, BVP features also have a great contribution in BP measurement. Through calculation, the cumulative importance of BVP features reaches 0.56 and 0.30 in the DBP and SBP RF regressors, respectively, verifying the effectiveness of RBP-CNN based BVP feature extraction.

# 5. Conclusion

This paper presents a video-based remote BP measurement scheme via convolutional network and RF feature fusion. We combine residual convolution, local and global attention mechanisms to design RBP-CNN for learning the implicit BP-related information in BVP spatially and temporally. Subsequently, we capture BMI, age, HR from facial video and analyze their correlation with BP. In this process, we find a strong correlation between DBP and SBP. At last, we use RF to fuse these features to achieve BP measurement and verify the rationality of our method by using the feature importance of RF. Our method achieves second place on the 3nd RePSS Challenge Track 2, and we believe we can do better in the future.

## 6. Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant62176124, Grant 62276135, and Grant 62361166670.

## References

- Lim, S. S., Vos, T., Flaxman, D. A., Danaei, G., Shibuya, K., Adair-Rohani, H., ... and Pelizzari, M. P. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet 380 (2012) 2224–2260.
- [2] Zhou, B., Perel, P., Mensah, G. A., and Ezzati, M, Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. Nature Reviews Cardiology 18 (2021) 785–802.
- [3] Olsen, H. M., Angell, Y. S., Asma, S., Boutouyrie, P., Burger, D., Chirinos, A. J., ... and Wang, G. J, A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on hypertension. The Lancet 388 (2016) 2665–2712.
- [4] Hassan, A. M., Malik, S. A., Fofi, D., Saad, N., Karasfi, B., Ali, S. Y., and Meriaudeau, F, Heart rate estimation using facial video: A review. Biomedical Signal Processing and Control 38 (2017) 346–360.
- [5] Rouast, V. P., Adam, T. M., Chiong, R., Cornforth, D., and Lux, E, Remote heart rate measurement using low-cost RGB face video: a technical literature review. Frontiers of Computer Science 12 (2018) 858–872.
- [6] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Wang, J. Z, Video-based heart rate measurement: Recent advances and future prospects. IEEE Transactions on Instrumentation and Measurement 68 (2018) 3600–3615.
- [7] Z. Yu, X. Li, and Zhao, G, Facial-video-based physiological signal measurement: Recent advances and affective applications. IEEE Signal Processing Magazine 38 (2021) 50–58
- [8] Xiao, H., Liu, T., Sun, Y., Li, Y., Zhao, S., and Avolio, A, Remote photoplethysmography for heart rate measurement: A review. Biomedical Signal Processing and Control 88 (2024) 105608.
- [9] Smith, P. R., J. Argod, Pépin, L. J., and Lévy, A. P. Pulse transit time: an appraisal of potential clinical applications. Thorax 54 (1999) 452–457.
- [10] D. Shao, Y. Yang, C. Liu, F. Tsow, H. Yu and N. Tao, Noncontact Monitoring Breathing Pattern, Exhalation Flow Rate and Pulse Transit Time. IEEE Transactions on Biomedical Engineering 61 (2014) 2760–2767.

- [11] Jeong, C. I., and Finkelstein, J. Introducing Contactless Blood Pressure Assessment Using a High Speed Video Camera. Journal of Medical Systems 40 (2016) 1–10.
- [12] Fan, X., Ye, Q., Yang, X., and Choudhury, D. S, Robust blood pressure estimation using an RGB camera. Journal of Ambient Intelligence and Humanized Computing 11 (2020) 4329–4336.
- [13] Wu, F. B., Wu, J. B., Tsai, R. B., and Hsu, P. C, A facial-image-based blood pressure measurement system without calibration. IEEE Transactions on Instrumentation and Measurement 71 (2022) 1–13.
- [14] Zhou, Y., Ni, H., Zhang, Q., and Wu, Q. The noninvasive blood pressure measurement based on facial images processing. IEEE Sensors Journal 19 (2019) 10624–10634.
- [15] Rong, M., and Li, K. A blood pressure prediction method based on imaging photoplethysmography in combination with machine learning. Biomedical Signal Processing and Control 64 (2021) 102328.
- [16] Luo, H., Yang, D., Barszczyk, A., Vempala, N., Wei, J., Wu, J. S., ... and Feng, P. Z, Smartphonebased blood pressure measurement using transdermal optical imaging technology. Circulation: Cardiovascular Imaging 12 (2019) e008857.
- [17] Poh, Z. M., McDuff, J. D., and Picard, W. R, Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE transactions on biomedical engineering 58 (2010) 7–11.
- [18] De Haan, G., and Jeanne, V, Robust pulse rate from chrominance-based rPPG. IEEE transactions on biomedical engineering 60 (2013) 2878–2886.
- [19] X. Li, J. Chen, G. Zhao and M. Pietikäinen, Remote Heart Rate Measurement from Face Videos under Realistic Situations, 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 4264–4271
- [20] Wang, W., Den Brinker, C. A., Stuijk, S., and De Haan, G. Algorithmic principles of remote PPG. IEEE Transactions on Biomedical Engineering 64 (2016) 1479–1491.
- [21] Casado, A. C., and López, B. M, Face2PPG: An unsupervised pipeline for blood volume pulse extraction from faces. IEEE Journal of Biomedical and Health Informatics 27 (2023) 5530–5541.
- [22] Chen, W., and McDuff, D, Deepphys: Video-based physiological measurement using convolutional attention networks. In Proceedings of the european conference on computer vision (ECCV), 2018, pp. 349–365.
- [23] Liu, X., Fromm, J., Patel, S., and McDuff, D, Multi-task temporal shift attention networks for on-device contactless vitals measurement. Advances in Neural Information Processing Systems 33 (2020) 19400–19411.
- [24] Liu, X., Hill, B., Jiang, Z., Patel, S., and McDuff, D, 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.
- [25] Yu, Z., Shen, Y., Shi, J., Zhao, H., Torr, H. P., and Zhao, G, 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.
- [26] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

- [27] Branco, P., Torgo, L., and Ribeiro, P. R, SMOGN: a pre-processing approach for imbalanced regression. In First international workshop on learning with imbalanced domains: Theory and applications. PMLR, 2017, pp. 36–50.
- [28] Steininger, M., Kobs, K., Davidson, P., Krause, A., and Hotho, A, Density-based weighting for imbalanced regression. Machine Learning 110 (2021) 2187–2211.
- [29] Yang, Y., Zha, K., Chen, Y., Wang, H., and Katabi, D, Delving into deep imbalanced regression. In International conference on machine learning. PMLR, 2021, pp. 11842–11851.
- [30] Ren, J., Zhang, M., Yu, C., and Liu, Z, Balanced mse for imbalanced visual regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 7926–7935.
- [31] Kuprashevich, M., and Tolstykh, I, Mivolo: Multi-input transformer for age and gender estimation. In International Conference on Analysis of Images, Social Networks and Texts. Cham: Springer Nature Switzerland, 2023, pp. 212–226.
- [32] Breiman, L, Random forests. Machine learning 45 (2001) 5-32.
- [33] McDuff, D, Camera Measurement of Physiological Vital Signs. ACM Computing Surveys 55 (2023) 1–40.
- [34] Li, X., Alikhani, I., Shi, J., Seppanen, T., Junttila, J., Majamaa-Voltti, K., ... and Zhao, G, 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 and gesture recognition, 2018, pp. 242–249.