PRV parameters extraction from rPPG signals in NIR images: adaptation and evaluation of NNs architectures designed for RGB

Marc Escrig-Villalonga
Digital Health Department
Instituto de Biomecánica de Valencia
València, Spain
marc.escrig@ibv.org

Ursula Martinez-Iranzo
Digital Health Department
Instituto de Biomecánica de Valencia
València, Spain
ursula.martinez@ibv.org

Cristina Herrera-Ligero
Digital Health Department
Instituto de Biomecánica de Valencia
València, Spain
cristina.herrera@ibv.org

Alberto Albiol
PRHLT Research Centre
Universidad Politécnica de València
València, Spain
marc.escrig@ibv.org

***Abstract*— Remote Photoplethysmography (rPPG) has emerged as a promising technique for assessing Pulse Rate Variability (PRV) using standard video cameras. While most existing research relies on RGB imaging and focuses on heart rate estimation, this work systematically evaluates the feasibility of extracting time-domain PRV parameters from near-infrared (NIR) images via deep learning. A proprietary dataset of 10 participants was recorded under controlled low-light conditions using an 850 nm light source and synchronized PPG signals as ground truth. Four state-of-the-art architectures—PhysNet, PhysFormer, EfficientPhys, and LSTCrPPG—were trained and evaluated using a random-search hyperparameter tuning framework. Results show that PhysNet consistently achieved the lowest Mean Absolute Error (MAE) when estimating both Mean of Normal-to-Normal Intervals (meanNNI) and Standard Deviation of Normal-to-Normal Intervals (SDNN), indicating superior performance in capturing average pulse intervals and short-term variability. However, each PRV metric required a different set of hyperparameters for the Physnet model, highlighting the importance of optimising model settings based on the specific physiological parameter being measured. While PhysFormer and LSTCrPPG are capable of learning complex spatiotemporal representations, they require larger datasets to avoid overfitting. In contrast, EfficientPhys proved less effective at preserving the morphological and dynamic features of the pulse waveform. These results underline the importance of adapting model design, loss functions and dataset design to reduce the error in PRV parameter estimation. Despite PhysNet's robust performance under data-constrained conditions, further refinements are required to reduce errors in short-term variability estimates. Additionally, exploring architectures that leverage pixel-intensity differences, alternative loss functions, and more sensitive imaging technologies (e.g., SWIR) hold potential to advance rPPG-based PRV monitoring in real-world applications.**

Keywords*— *Pulse Rate Variability (PRV), Remote Photoplethysmography (rPPG), Near Infrared Image (NIR), Deep Learning, Driver Monitoring, Health Monitoring, Physiological Signal Processing

I. INTRODUCTION

Monitoring physiological states is essential not only in clinical settings but also in activities that require sustained attention, such as driving. Among the various physiological markers, Pulse Rate Variability (PRV) has emerged as a key indicator for assessing stress, fatigue, or cognitive load [1]. Conventional Photoplethysmography (PPG) techniques, while tested and reliable, often rely on contact-based sensors that may be impractical in real-world scenarios.

In this context, remote Photoplethysmography (rPPG) has gained attention as a non-invasive alternative capable of detecting subtle skin changes through standard video cameras, offering the possibility to measure PRV without the need for contact devices. This technology represents a significant step forward in physiological monitoring by enabling continuous, contact-free measurements in situations where traditional sensors may not be suitable.

Current rPPG signal extraction methodologies are generally classified into two main groups: classical signal and image processing methods, and deep learning-based approaches [2]. Although both have shown promising results in controlled conditions, deep learning-based solutions have demonstrated greater robustness to lighting and motion variations, enhancing the accuracy of rPPG signal extraction under non-controlled conditions [2].

Most of the available research employs RGB cameras for rPPG estimation, whereas near-infrared (NIR) technology has been less explored [3]. rPPG extraction is generally more robust at visible wavelengths —especially in the green region of the spectrum, where light absorption by the skin is maximal —than in the NIR region. In NIR imaging, differences in light attenuation and absorption can reduce pulse amplitude and limit the available signal range. Moreover, the lower surface reflectance in the NIR spectrum necessitates stable, high-quality infrared illumination to maintain a sufficient signal-to-noise ratio. Consequently, processing models and deep learning networks must be specifically adapted to the unique physiological and photometric characteristics inherent to NIR capture.

Although NIR cameras offer significant benefits, such as reduced sensitivity to solar radiation in environments with varying lighting and improved performance under low-light conditions through the use of dedicated NIR illumination (which can be added without affecting the subject's vision), most studies using NIR focus solely on average heart rate estimation. As a result, the accuracy of other PRV parameters derived from rPPG NIR remains insufficiently investigated [4]. This research gap underscores the need for a

comprehensive evaluation of PRV extraction methods to develop robust, real-world monitoring systems.

To the best of our knowledge, this study is the first to systematically compare PRV parameters extraction from rPPG signals derived from NIR images using multiple deep learning models, including PhysNet, PhysFormer, EfficientPhys, and LSTCrPPG. These methods were selected based on their relevance in the literature and architectural differences, which allows for an assessment of how distinct approaches influence PRV from rPPG NIR estimation accuracy. Using a proprietary dataset, we evaluate the performance of these models under low-light conditions. For this analysis, only models that take image sequences as inputs have been considered, excluding those relying on intensity differences between pixels in consecutive frames, which will be addressed in future research.

This work assesses the feasibility of deep learning models for PRV parameters estimation via rPPG using NIR imaging, offering key insights into their robustness and reliability under variable lighting conditions. By highlighting both the strengths and limitations of these methods, this study contributes to the development of more efficient and adaptable monitoring systems for diverse real-world applications.

II. Methodology

A. Dataset description

A dataset of 10 subjects (4 males and 6 females) was acquired. Participants' age ranged from 35 to 49 years with no reported cardiovascular or dermatological issues. Two separate measurements were recorded on different days to ensure day-to-day variability was captured, resulting in a total of 20 recordings. The subject remained in an upright sitting posture and instrumented in a darkened room throughout the sessions, with the camera positioned in front of them, at approximately 1.5 meters, and angled to capture the face. The forehead, face and lower neck were exposed to the cameras. Subjects were required to remain still and quiet, breathing normally.

During each recording session, a NIR camera [5] was used to capture images synchronously with a resolution of 640x640 pixels at a sampling rate of 60 Hz (see Figure 1 for the experimental setup). Each session lasted approximately 1 minute, resulting in ~3600 frames. In total, around 72,000 images were collected. Image capture was synchronised with the acquisition of PPG signal. For this purpose, the BiosignalsPlux acquisition system [6] was used together with a PPG sensor [7] placed on the index finger of the left hand.

Fig. 1. Experimental setup and example of NIR imaging. The left image shows the experimental setup, including the NIR camera, NIR illumination source, and PPG sensor placed on the subject's index finger. The right image presents an example of a NIR facial image captured during the experiment, demonstrating the imaging conditions and illumination uniformity.

The recording environment was illuminated exclusively by an 850 nm light source. To ensure that this NIR source was the primary illumination, all visible artificial and natural ambient light sources were eliminated, minimizing potential interference. Additionally, the light source was equipped with an electronic system and a light diffusion lens to provide continuous, stable, and uniform illumination throughout the recordings.

B. Dataset preprocess

Preprocessing was applied to both the ground-truth PPG signals and the NIR images. For the PPG signals, a baseline drift removal was performed using the detrend function from SciPy [8], followed by a second-order Butterworth bandpass filter (0.33–3.5 Hz), which also serves as an anti-aliasing measure. The signals, initially sampled at 1000 Hz, were downsampled to 60 Hz to match the camera's frame rate. Finally, each PPG signal was normalized per clip by subtracting its median and dividing by the median absolute deviation (MAD). The resulting signal was then scaled to the range [–1, 1] by dividing by its maximum absolute value. This robust normalization is more resistant to outliers than standard z-scoring and ensures consistent amplitude scaling across sequences.

For NIR images, face detection was performed using the MediaPipe Facemesh model [9], which locates and tracks the subject's face. Minor deviations in the detected position were corrected by computing an average face location over the entire recording, minimizing tracking drift. This was achieved by averaging all bounding boxes detected by MediaPipe Facemesh frame-by-frame, generating a stable reference position for cropping. The cropped images were then resized to 128×128 pixels to provide a uniform input format across all models.

C. Architectures and experimental configuration

This study focuses on neural network architectures that process variable-length sequences of NIR images as inputs and generate a continuous rPPG signal as an output through a regression task (see Figure 2 for an overview of the complete pipeline). Four distinct models (PhysNet, PhysFormer, EfficientPhys y LSTCrPPG) were selected. Table I summarizes the main features of the models. All of them were originally designed for three-channel RGB inputs and subsequently modified to accept single-channel NIR images. Additionally, the input was adapted to accommodate 128×128 images and variable-length image sequences. The RemoteBiosensing framework [10] was adapted and employed to train and evaluate the models.

TABLE I. Models Description

Model	Main Features
PhysNet [11]	Utilizes 3D convolutions to extract semantic rPPG features in both spatial and temporal domains, enabling robust contextual representations and reducing temporal fluctuations. Its encoder-decoder architecture minimizes redundancy and temporal noise.
PhysFormer [12]	Adaptively integrates local and global spatiotemporal features by capturing temporal differences in skin color to enhance prediction accuracy. Its architecture comprises a shallow stem with convolutional blocks, a tube tokenizer, multiple temporal difference transformer blocks, and an rPPG predictor head.

Model	Main Features
EfficientPhys [13]	A convolution-based architecture known for its simplicity and speed. It operates as a single-branch network that processes raw video frames without preprocessing to generate a first-derivative PPG signal. Key components include a custom normalization layer (using temporal differences and batch normalization), self-attention modules, and tensor shift modules (TSM) for efficient spatiotemporal
LSTCrPPG [14]	Employs an hourglass-shaped 3D CNN with skip connections to capture both short- and long-term spatiotemporal features. It incorporates a Temporal Attention Refinement Module (TARM) to align these features and utilizes a hybrid loss—combining time-domain and frequency-domain components—for enhanced rPPG signal estimation.

An 80/20 split of the dataset was used for training and validation/testing, ensuring that subjects in the validation/test set were not included in the training phase. To optimize performance, a random-search hyperparameter tuning process was implemented. The hyperparameter space encompassed the following parameters: learning rate (LR) ∈ {0.05, 0.01, 0.005}; input image sequence length (Len) ∈ {120, 240, 300}; batch size (BS) ∈ {4, 8, 16}; overlap in input image sequence (Ov) ∈ {0, 60}. Three different optimization algorithms (Opt) were evaluated: AdamW (Adaptive Moment Estimation with decoupled weight decay), SGD (Stochastic Gradient Descent), and RAdam (Rectified Adam optimizer). The loss function was selected from one of the following: negative Pearson correlation loss (neg_pearson), Fourier transform-based loss (FTT loss), LSTCrPPG loss, and blood volume pulse velocity loss (BVPVelocityLoss). The LSTCrPPGLoss [14] combines time-domain mean square error with a frequency-domain loss computed from the log power spectral densities of the signals, ensuring alignment in both temporal and spectral features, while the BVPVelocityLoss enhances waveform fidelity by integrating negative Pearson correlation with additional penalties based on peak alignment and signal derivative differences. Each model was trained for 15 epochs per hyperparameter configuration. A total of 150 different configurations were executed per architecture.

After training, the output signals were refined and directly compared with the ground-truth signals to evaluate performance (also illustrated in Figure 2). Table II and III details the hyperparameter combinations that yielded the best metrics for each architecture, thereby highlighting the influence of different design choices on the extraction of reliable signals from NIR images.

D. rPPG signal refinement and IBIs extraction

The same detrending, band-pass filtering, and normalization steps used for the ground truth PPG signal were also performed on the model's output rPPG signal to ensure a fair comparison.

After that, the rPPG signal is processed to extract inter-beat intervals (IBIs) by identifying systolic peaks corresponding to each pulse. Peak detection was performed using a custom algorithm, which identifies local maxima by comparing each sample with its neighbours within defined segments. This method integrates an adjustable quality criterion to validate detected peaks, ensuring that only physiologically relevant maxima are retained while suppressing spurious detections. Additionally, an adaptive thresholding mechanism was applied to dynamically adjust detection sensitivity based on signal characteristics. To enhance robustness, the detection function operates in a 10-second sliding window (with a 5-second overlap) to reduce the influence of high-amplitude artifacts that could obscure true systolic peaks.

To ensure the physiological plausibility of the extracted IBIs, outliers corresponding to heart rates outside the physiological range (30–200 beats per minute) were first removed. An iterative filtering approach was then applied to refine the detected peaks and validate the remaining intervals. The acceptable IBI range was dynamically recalculated as the rolling mean ± k·SD of the most recent valid intervals, where k controls the sensitivity of the threshold. Any intervals exceeding these adaptive bounds were excluded, and the process was repeated until no further outliers were detected. Finally, linear interpolation was used to reconstruct a continuous IBI sequence while preserving physiological consistency.

An identical peak detection and iterative filtering procedure was applied to the ground-truth PPG signal to ensure direct comparability with the rPPG-derived IBIs.

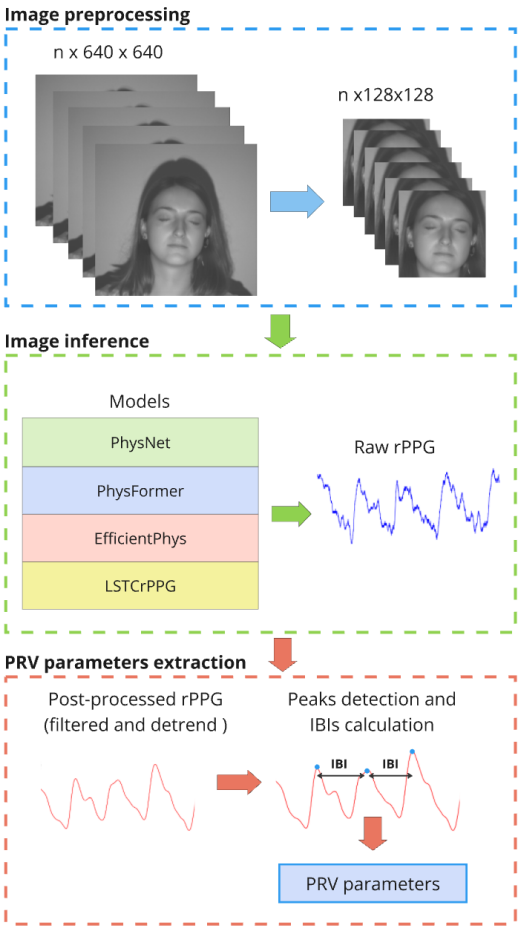

Fig. 2. Overview of the proposed rPPG pipeline for PRV estimation. In the first blue stage, NIR images are preprocessed—this includes face detection, cropping, and resizing to 128×128. The processed frames are then passed to a deep learning models which output the raw rPPG signal (green stage). A post-processing step (red stage) detects systolic peaks, computes inter-beat intervals (IBIs), and derives the final time-domain PRV parameters (e.g., meanNNI and SDNN).

E. Evaluation Metrics

This study focuses on time-domain PRV parameters, which are widely used in short-term recordings due to their reliability and lower sensitivity to recording duration constraints compared to frequency-domain or nonlinear metrics [15]. Among these parameters, the two most informative indices were selected for their robustness: Mean of Normal-to-Normal Intervals (meanNNI), which represents the central tendency of IBIs, and Standard Deviation of Normal-to-Normal Intervals (SDNN), which quantifies short-term heart rate variability.

To evaluate model performance, Mean Absolute Error (MAE) was used as the primary metric due to its interpretability and robustness against outliers, offering a more balanced measure of average deviation than Root Mean Square Error (RMSE), which is more sensitive to extreme values.

III. Results and Discussion

Tables II and III present the MAE for meanNNI and SDNN, respectively. PhysNet achieved the lowest MAE for both metrics, indicating superior performance in capturing both the central tendency (meanNNI) and variability (SDNN) of pulse rate.

TABLE II. MAE meanNNI (ms) Across Different Models and Hyperparameter Configurations

Model	MAE MeanNNI (ms)	Hyperparameters
PhysNet	68.8	LR=0.001, Len=120, BS=4, Ov=0, Opt=AdamW, Loss=neg_pearson, Training Samples=57 600
PhysFormer	83.0	LR=0.01, Len=300, BS=4, Ov=0, Opt=AdamW, Loss=neg_pearson/FTT, Training Samples=57 600
EfficientPhys	123.1	LR=0.01, Len=120, BS=16, Ov=0, Opt=AdamW, Loss=FTT, Training Samples=57 600
LSTCrPPG	99.3	LR=0.05, Len=120, BS=4, Ov=0, Opt=AdamW, Loss=BVPVelocityLoss, Training Samples=57 600

TABLE III. MAE SDNN (ms) Across Different Models and Hyperparameter Configurations

Model	MAE SDNN (ms)	Hyperparameters
PhysNet	103.75	LR=0.005, Len=300, BS=4, Ov=60, Opt=RAdam, Loss=BVPVelocityLoss, Training Samples=67 200
PhysFormer	202.4	LR=0.01, Len=300, BS=4, Ov=0, Opt=AdamW, Loss=neg_pearson/FTT, Training Samples=57 600
EfficientPhys	263.8	LR=0.01, Len=120, BS=8, Ov=60, Opt=RAdam, Loss=neg_pearson, Training Samples=67 200
LSTCrPPG	261.31	LR=0.01, Len=120, BS=4, Ov=0, Opt=RAdam, Loss=neg_pearson, Training Samples=57 600

A. MAE for MeanNNI and SDNN

For meanNNI, PhysNet obtained a MAE of 68.8 ms, outperforming PhysFormer (92.0 ms) and notably exceeding EfficientPhys (117.2 ms) and LSTCrPPG (117.0 ms). A closer look at the hyperparameters shows that PhysNet's best configuration utilized a smaller input length (120 frames → 2 s), no overlap, and AdamW optimizer with a negative Pearson correlation loss. A similar trend was observed for SDNN, where PhysNet again recorded the lowest MAE (103.75 ms). Interestingly, in this case, the best-performing hyperparameters for PhysNet included a larger input length (300 frames → 5 s), an overlap of 60 frames, and a different optimizer (RAdam) with the BVPVelocityLoss function. These quantitative results are further illustrated in Figure 3, which shows an example segment from the test dataset comparing the ground-truth PPG (blue) with PhysNet's raw (orange) and filtered outputs (green).

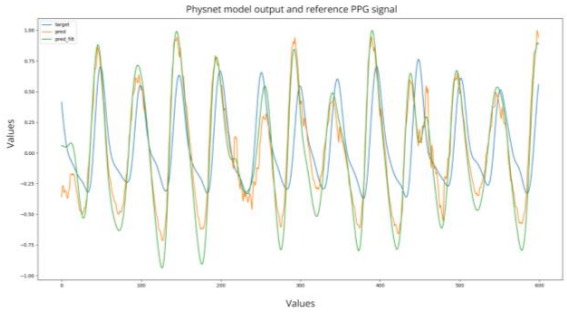

Fig. 3. Example segment from the test dataset illustrating PhysNet's performance. The ground-truth PPG (blue) is compared to PhysNet's raw output (orange) and its filtered version (green), demonstrating the model's ability to approximate the waveform morphology and the additional refinement provided by post-processing.

This result implies that, while meanNNI can be estimated accurately with shorter sequence lengths and simpler optimization strategies, characterizing SDNN might benefit from longer temporal windows. With short windows, in addition to fewer valid intervals, there is greater sensitivity to artifacts and outliers that can distort the true variability. Consequently, the models may end up overestimating or underestimating the variability, which increases the MAE in the SDNN estimation. The differences in optimal configurations for meanNNI and SDNN further indicate that no single set of parameters in Physnet necessarily yields the best performance across all PRV metrics, emphasizing the need for tailored approaches depending on the specific physiological feature of interest.

In contrast, PhysFormer, EfficientPhys, and LSTCrPPG exhibited higher MAEs across both meanNNI and SDNN. Although PhysFormer applied a single configuration for both metrics (learning rate=0.01, input length=300, batch size=4, overlap=0, AdamW, neg_pearson/ftt), it did not adequately capture short-term variability. EfficientPhys (123.1 ms meanNNI; 263.8 ms SDNN) and LSTCrPPG (99.3 ms meanNNI; 261.31 ms SDNN) likewise fell short, suggesting that their respective architectures and hyperparameter choices may be less suited to the current NIR-based dataset.

B. Architecture-Specific Insights

A detailed comparison of PhysNet, PhysFormer, LSTCrPPG, and EfficientPhys under small NIR dataset conditions reveals distinct performance profiles. PhysNet, leveraging 3D convolutions, can be tuned to prioritize beat interval consistency (improving meanNNI) or short-term fluctuation capture (improving SDNN). However, no single configuration has excelled at both objectives simultaneously.

PhysFormer (transformer-based) and LSTCrPPG (hourglass 3D CNN) capture both beat intervals and subtle variability but exhibit rapid overfitting, suggesting a larger dataset could better exploit their complex spatiotemporal architectures. In contrast, EfficientPhys, with its simpler single-branch design, struggled to accurately reproduce the pulse waveform under these conditions, indicating limited representational capacity for subtle NIR signal variations

Overall, these findings underscore the importance of aligning model capacity, dataset size, and loss functions with the specific target (morphology vs. dynamics) in PRV from rPPG NIR estimation. While PhysNet proved most adaptable under data constraints, it lacks the ability to capture both waveform shape and fluctuation simultaneously at optimal levels. PhysFormer and LSTCrPPG may excel given a more extensive training corpus, whereas EfficientPhys appears inherently limited for tasks requiring nuanced waveform representation.

C. Real-World Implication

From a practical standpoint, PhysNet's superior performance in estimating both meanNNI and SDNN underscores its potential suitability for real-world monitoring scenarios. In particular, applications primarily concerned with averaged PRV values (e.g., meanNNI) may benefit from PhysNet's relatively low error rates, enabling reliable trend analysis over time. However, the SDNN errors observed across all models—even for PhysNet—remain too large for applications requiring fine-grained insights into mental states, where subtle changes on the order of 20–30 ms can be critical.

D. Limitations and Future Directions

Despite the promising results, this study is subject to several limitations. First, the dataset size remains small (10 subjects), reducing the models' capacity to generalize across broader demographic and physiological variations. Additionally, the short recording durations (~1 minute per session) and tightly controlled lighting conditions may not fully capture real-world scenarios where ambient illumination, motion artifacts, and subject-specific factors can significantly affect rPPG signal quality. Moreover, while PhysNet demonstrated stronger adaptability under these constraints, even its estimated SDNN values remain too coarse for applications requiring highly precise measures of autonomic function.

Future efforts will explore architectures that process pixel-intensity differences between consecutive frames, potentially capturing subtle blood flow variations more effectively. Investigating alternative loss functions that prioritize peak localization could also yield more accurate beat detection for PRV analysis. Building on the 3D convolution, transformer, and hourglass designs tested here, refined architectures may further enhance performance under limited data conditions. Finally, Short-Wave Infrared (SWIR) cameras, offering higher sensitivity in the IR range, represent a promising avenue for improving signal quality in challenging lighting environments, thereby moving closer to robust, real-world rPPG implementations.

ACKNOWLEDGMENT

This work has been funded by MEDUSA project. MEDUSA is a Technological Network of applied engineering to the development of intelligent solutions for **human-centered autonomous driving** (CER-20231011), Red de Excelencia CERVERA (CERVERA Network of Excellence) financed by the Ministerio de Ciencia, Innovación y Universidades through the Centro para el Desarrollo Tecnológico y la Innovación E.P.E. (CDTI), charged to the European Union Recovery and Resilience Mechanism.

The author(s) gratefully acknowledges the computer resources at Artemisa, funded by the European Union ERDF and Comunitat Valenciana as well as the technical support provided by the Instituto de Física Corpuscular, IFIC (CSIC-UV).

REFERENCES

[1] A. Burlacu, C. Brinza, A. Brezulianu, and A. Covic, 'Accurate and early detection of sleepiness, fatigue and stress levels in drivers through Heart Rate Variability parameters: a systematic review', *Rev Cardiovasc Med*, vol. 22, no. 3, pp. 845–852, Sep. 2021, doi: 10.31083/j.rcm2203090.

[2] H. Xiao, T. Liu, Y. Sun, Y. Li, S. Zhao, and A. Avolio, 'Remote photoplethysmography for heart rate measurement: A review', *Biomedical Signal Processing and Control*, vol. 88, p. 105608, Feb. 2024, doi: 10.1016/j.bspc.2023.105608.

[3] Chun-Hong Cheng *et al.*, 'Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda.', *Sensors*, vol. 21, no. 18, p. 6296, Sep. 2021, doi: 10.3390/s21186296.

[4] J. Wang, C. Shan, L. Liu, and Z. Hou, 'Camera-based physiological measurement: Recent advances and future prospects', *Neurocomputing*, vol. 575, p. 127282, Mar. 2024, doi: 10.1016/j.neucom.2024.127282.

[5] 'GS3-U3-41C6NIR-C: 4.1 MP, 90 FPS, CMOSIS CMV4000-3E12, NIR | Teledyne Vision Solutions'. Accessed: Mar. 2, 2025. [Online]. Available: https://www.teledynevisionsolutions.com/products/grasshopper3-usb3?vertical=machine+vision&segment=iis

[6] '8-Channel biosignalsplux Kit', PLUX Biosignals. Accessed: Mar. 2, 2025. [Online]. Available: https://www.pluxbiosignals.com/products/8-channel-biosignals-kit

[7] 'PLUX Biosignals | Blood Volume Pulse (BVP) Finger Clip Sensor'. Accessed: Mar. 2, 2025. [Online]. Available: https://www.pluxbiosignals.com/products/blood-volume-pulse-bvp-sensor

[8] 'detrend — SciPy v1.14.1 Manual'. Accessed: Mar. 2, 2025. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.detrend.html

[9] 'mediapipe/docs/solutions/face_mesh.md at master · google-ai-edge/mediapipe', GitHub. Accessed: Mar. 2, 2025. [Online]. Available: https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/face_mesh.md

[10] D.-Y. Kim *et al.*, 'Remote Bio-Sensing: Open Source Benchmark Framework for Fair Evaluation of rPPG', Aug. 18, 2023, *arXiv*: arXiv:2307.12644. doi: 10.48550/arXiv.2307.12644.

[11] Z. Yu, X. Li, and G. Zhao, 'Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks', Jul. 31, 2019, *arXiv*: arXiv:1905.02419. doi: 10.48550/arXiv.1905.02419.

[12] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip H. S. Torr, and Guoying Zhao, 'PhysFormer: Facial Video-based Physiological Measurement with Temporal Difference Transformer', *Computer Vision and Pattern Recognition*, 2021, doi: 10.1109/cvpr52688.2022.00415.

[13] Xin Liu, B. Hill, Ziheng Jiang, Shwetak N. Patel, and Daniel J. McDuff, 'EfficientPhys: Enabling Simple, Fast and Accurate Camera-Based Vitals Measurement', *ArXiv*, 2021.

[14] J. S. Lee, G. Hwang, M. Ryu, and S. J. Lee, 'LSTC-rPPG: Long Short-Term Convolutional Network for Remote Photoplethysmography', in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2023, pp. 6015–6023. doi: 10.1109/CVPRW59228.2023.00640.

[15] F. Shaffer, Z. M. Meehan, and C. L. Zerr, 'A Critical Review of Ultra-Short-Term Heart Rate Variability Norms Research', *Front. Neurosci.*, vol. 14, Nov. 2020, doi: 10.3389/fnins.2020.594880.