

Review

Towards Objective Emotional Monitoring in Children with Cerebral Palsy: A Review of rPPG and Multimodal Approaches

Martha Xóchitl Nava-Bautista ¹, Víctor H. Castillo-Topete ¹, Alberto J. Molina-Cantero ²
and Isabel M. Gómez-González ^{2,*}

¹ Escuela de Ingeniería Mecánica y Eléctrica, Universidad de Colima, Coquimatlán 28400, Mexico; xnava@ucol.mx (M.X.N.-B.); victorc@ucol.mx (V.H.C.-T.)

² Departamento de Tecnología Electrónica, E.T.S.I. Informática, Universidad de Sevilla, 41012 Sevilla, Spain; almolina@us.es

* Correspondence: igomez@us.es

Abstract

Non-contact physiological monitoring based on remote PPG (rPPG) offers a viable alternative for the care of pediatric populations, particularly for children with cerebral palsy (CP) who present unique communication and mobility challenges. This paper presents a review of the literature on the use of rPPG for the estimation of vital signs and its application in emotional monitoring. Following the PRISMA 2020 guidelines as a methodological framework for searching and filtering, an exhaustive search was conducted in the IEEE Xplore and Scopus databases covering the period from 2017 to 2024. A total of 35 studies were selected for analysis. The review examines the evolution of rPPG algorithms—from classical mathematical approaches to recent deep-learning-based architectures—identifying critical technical challenges such as motion artifacts caused by spasticity and variations in lighting conditions. The results reveal that while rPPG has reached technical maturity for monitoring core physiological parameters such as heart rate, its application to robust emotion detection in children with CP remains limited. The main limitation identified across the surveyed literature is the critical scarcity of public or clinical datasets featuring pediatric CP cohorts. Finally, the potential of multimodal integration—combining rPPG with eye-tracking and wearable sensors—is discussed as a promising pathway toward objective emotional monitoring. Such an approach could enhance communication, support rehabilitation processes, and ultimately improve the quality of life of children with cerebral palsy and their caregivers.

Keywords: assistive technology; cerebral palsy; emotion recognition; heart rate estimation; multimodal monitoring; rPPG



Academic Editors: Adrian Gligor and Cristina Veres

Received: 15 April 2026

Revised: 22 May 2026

Accepted: 27 May 2026

Published: 1 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Physiological and emotional monitoring in children with cerebral palsy (CP) is necessary for providing comprehensive care. This condition presents unique challenges due to limitations in mobility and communication. Therefore, complicate the objective interpretation of a child's affective state and well-being by caregivers and healthcare professionals. Hence, developing non-invasive technological tools is paramount to enable continuous monitoring without increasing patient stress or discomfort.

Remote PPG (rPPG) has emerged as a promising technology for this purpose. This technique allows for the estimation of vital signs, such as Heart Rate (HR), by analyzing

skin color changes caused by blood volume variations, captured through conventional or infrared video cameras [1–3]. Unlike traditional contact sensors, rPPG offers a contactless alternative that is particularly valuable in environments where it is difficult to use external devices that require physical contact [4].

The technical development of rPPG has progressed from classical signal processing algorithms to advanced architectures. Recent research emphasizes the implementation of Field Programmable Gate Array (FPGA)-based hardware accelerators that utilize Independent Component Analysis (ICA) to achieve real-time HR detection with high computational speed and precision [1,5]. In parallel, the rise of deep learning has enabled the development of adaptive weight networks, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM), that optimize the quality of signals extracted from multiple facial Regions Of Interest (ROIs), facilitating their implementation even on mobile devices and embedded systems [6,7]. ROI selection is critical for achieving more accurate results in rPPG, and universal ROIs can be proposed for use in various scenarios [8].

However, the application of rPPG in children with CP involves critical technical challenges. Motion artifacts caused by spasticity and involuntary movements, along with variations in lighting conditions, can significantly degrade the signal. For this reason, recent architectures implement adaptive-weight networks across multiple facial regions to extract stable physiological features under dynamic noise [7]. Mitigating these issues involves integrating Kalman filters with pose constraints and deep facial trackers (Deep ROI Tracker) for maintaining precision in skin tracking [2]. Similarly, environmental noise reduction techniques and the use of specific color channels, such as the blue channel in smartphone recordings, have proven effective in enhancing the robustness of measurements [9].

Although rPPG has reached a notable level of maturity for physiological parameter monitoring, its specific application for emotion detection in children with CP remains limited. The literature suggests that the path toward objective monitoring lies in multimodal integration [10].

Beyond traditional rPPG methods, several emerging optical technologies have demonstrated significant potential for non-invasive monitoring of cardiovascular parameters. For instance, Diffuse Speckle Pulsatile Flowmetry (DSPF) has recently been introduced as a standalone, compact system for simultaneous blood flow and volume monitoring. Unlike standard imaging techniques, DSPF leverages laser speckle contrast analysis to achieve deep tissue penetration and high-frequency measurements, proving comparable to gold-standard Doppler ultrasound in assessing endothelial function. Other advancements include novel optical approaches for heart rate and blood flow measurement reported in recent studies, such as specialized bio-engineering frameworks [11]. Integrating these technologies as complementary information to rPPG provides a more comprehensive landscape of the state of the art in optical physiological monitoring.

This article presents a comprehensive literature review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines, analyzing studies published in the IEEE Xplore and Scopus databases between 2017 and 2024. The transition from traditional methods toward deep learning approaches is examined, technical gaps in CP are identified, and the potential of multimodal systems as a promising avenue for improving rehabilitation and the quality of life for both children and their caregivers is discussed.

This work is structured as follows: Section 2 details the search methods and strategies, including the databases consulted and the PRISMA selection criteria. Next, Section 3 addresses the theoretical foundations of rPPG, the evolution of algorithms, and technical challenges. Section 4 then examines the applications of rPPG in physiological monitoring.

Section 5 focuses on emotion recognition in children with CP, analyzing its specific challenges and limitations. Subsequently, Section 6 discusses the difficulties encountered and future perspectives, such as the combination of rPPG with other techniques. Finally, Section 7 presents the conclusions and methodological recommendations for future research in pediatric populations with CP.

2. Materials and Methods

Following the PRISMA 2020 guidelines, a structured literature review protocol was designed to identify, select and analyze the literature concerning rPPG and its application in monitoring children with CP. A completed PRISMA 2020 checklist is provided in the Supplementary Materials.

2.1. Databases and Sources

The literature search was conducted in IEEE Xplore, and Scopus. These databases were selected capturing the most recent advancements in rPPG algorithms. Capturing recent state-of-the-art developments, we supplemented this search with manual backward snowballing, screening the bibliographies of the primary studies to identify further relevant works meeting our inclusion criteria.

Additional references were identified through manual searches in the bibliographies of relevant studies.

2.2. Search Strategy

The literature search was executed in two stages across the selected databases. On 12 September 2024, the search was conducted in IEEE Xplore. The first query utilized the phrase “heart rate estimation by video analysis”, restricted to journal articles published between 2017 and 2024. A second query was performed using the terms “Emotion Recognition” AND “cerebral palsy”; due to the specificity of the topic and the limited number of results, filters were expanded to include both journals and conference proceedings for the same period.

On 18 September 2024, the process was replicated in Scopus. Following the same protocol, the initial search for “heart rate estimation by video analysis” was filtered for journal articles (2017–2024). Subsequently, the search was expanded to include “Emotion Recognition” AND “cerebral palsy” across journals and conferences, broadly encompassing the state of the art.

Manual backward reference searching (snowballing) was subsequently conducted on the bibliographies of the primary studies to identify further relevant literature. Full-text articles were excluded if they did not provide original empirical data or if their technical focus deviated from the core objectives of rPPG and CP monitoring.

The search focused on relevant articles in English, containing the selected keywords in the title and abstract, and published between 2017 and 2024. A total of 225 records were initially identified (78 from IEEE Xplore and 147 from Scopus). After removing duplicates and records with incomplete metadata, 190 unique records remained for the screening phase.

2.3. Inclusion and Exclusion Criteria

Maintaining the quality and relevance of the findings, the following criteria were enforced:

Inclusion:

- Studies published in peer-reviewed journals and full conference papers.
- Articles in English (and [Spanish, if applicable]).

- Relevant articles reporting experiments, analyzes, and/or evaluations of the proposed approach.
- Studies involving pediatric populations, children with disabilities, or scenarios simulating high-motion interference (relevant for CP).

Exclusion:

- Editorials, conference abstracts without full papers and opinion papers.
- Articles that do not report experiments, analyses, and/or evaluations of the proposed approach.
- Articles that lack empirical support through evaluation or analysis.

2.4. Screening and Selection

The selection process followed a multi-stage approach. Initially, duplicate records were removed. Titles and abstracts were screened independently by two reviewers. Full texts were assessed against the inclusion criteria. Disagreements were resolved by discussion or by consulting a third reviewer. The selection process is illustrated in a PRISMA flow diagram (Figure 1). Initially, 190 records were screened based on title and abstract. From these, 76 records were excluded for not meeting the pre-defined inclusion criteria, primarily due to being published in languages other than English or not being relevant to the clinical scope. Consequently, 114 reports were sought for retrieval, of which 12 reports could not be retrieved (e.g., lack of full-text access), leaving 102 reports to be assessed for eligibility through full-text review.

During the eligibility phase, 67 reports were excluded for the following reasons:

- Editorials, opinion papers, and conference abstracts without a corresponding full paper ($n = 19$).
- Articles that did not report experiments, technical analyses, or evaluations of the proposed approach ($n = 29$).
- Articles lacking empirical support through formal evaluation or rigorous analysis ($n = 19$).

After this systematic filtering, 35 studies were included for the final quality assessment and synthesis.

2.5. Data Extraction and Synthesis

The questions used for the analysis were as follows:

- QA1: Does it include experiments derived from the proposal?
- QA2: Does it include relevant or sufficient articles on the subject?
- QA3: Does it include an analysis and/or evaluation of the proposal, or does it provide quantitative results or an exhaustive analysis?
- QA4: Does it utilize experimental protocols?
- QA5: Does it incorporate elements for the protection of sensitive data?

The responses are categorized as follows: Y = Yes (value: 1), N = No (value: 0), and P = Partial/Few (value: 0.5). Only studies exceeding a cumulative threshold of 3.0 were considered for the final synthesis. This rigorous quality assessment ensured that only 18.4% of the initially screened records (35 out of 190) were included, reflecting a high specificity in the selection process; while the topic is highly specialized, this final selection represents the most methodologically sound and relevant contributions to the field of rPPG for CP published between 2017 and 2024.

The validation results are presented in (Table 1) Quality Assessment. For each included study, the following information was extracted:

- Study reference (title, authors and keywords).

- Methodological data: algorithm used (e.g., CHROM, POS, ICA, CNN, Masked Autoencoder).
- Practical scope, specifying applications and key concepts.
- Empirical evidence, highlighting the key results.
- Population and context: sample size, age group (pediatric vs. adult), and lighting/motion conditions.

The extracted information was synthesized in a narrative form, complemented with summary tables and contrasting methodologies, outcomes, and limitations across studies. The results are presented in Table 2 (techniques, algorithms and applications) and in Table 3 (key concepts and key results), and additional information is provided in Table 4.

Table 1. Quality assessment.

Reference	QA1	QA2	QA3	QA4	QA5	Total Value
[1]	1	1	1	1	1	5
[2]	1	1	1	1	0	4
[4]	1	0.5	1	1	0	3.5
[7]	1	1	1	1	0	4
[9]	1	1	1	1	0	4
[10]	1	1	1	1	1	5
[12]	1	1	1	1	0.5	4.5
[13]	1	1	1	1	0	4
[14]	0	1	1	1	0	3
[15]	1	1	1	1	0	4
[16]	1	1	1	1	1	5
[17]	1	1	1	1	0.5	4.5
[18]	1	1	1	1	1	5
[19]	1	1	1	1	0.5	4.5
[20]	1	1	1	1	0	4
[21]	1	1	1	1	0	4
[5]	1	1	1	1	0	4
[22]	1	1	1	1	0	4
[6]	1	1	1	1	1	5
[3]	1	1	1	1	1	5
[23]	1	1	1	1	0	4
[24]	1	1	1	1	1	5
[25]	1	1	1	1	0.5	4.5
[8]	1	1	1	1	0	4
[26]	1	1	1	1	0	4
[27]	1	1	1	1	0	4
[28]	1	1	1	1	0	4
[29]	1	1	1	1	0.5	4.5

Table 2. Techniques, algorithms and applications.

Ref	Techniques/Algorithms	Applications
[12]	rPPG-MAE: Masked Autoencoder, Vision Transformer (ViT), PC-STMap augmentation	Multimodal estimation: HR, HRV, and RF
[30]	TranPulse; Video Swin Transformer (3D); Asymmetric Encoder-Decoder	HR estimation; Multi-physiological signal disentanglement
[28]	pyVHR framework: POS, CHROM, PCA, SSR, ICA, GREEN, LGI, PBV, MTTs-CAN	HR estimation, HRV, blood volumetric pulse (BVP) estimation.
[16]	Cascade residual CNN-FPNR, ICNet, PCA, FFT, SVM classification, GEVD	Multimodal Estimation: HR, Respiratory Rate (RR), SpO ₂ , and Temperature.
[19]	BlazeFace, MediaPipe, ICA, Detrending Filter, ResNet blocks	Multimodal Estimation: HR, HRV, SpO ₂ , and Blood Pressure (BP)
[2]	DFT-KF-POS: ECO tracker, PFLD landmark detector, Pose Constrained Kalman Filter, POS + BCG fusion via FIR notch filters	Multimodal HR Estimation (BCG + rPPG)
[3]	CPulse: Homomorphic filtering, Empirical Wavelet Transform (EWT), and Principal Component Analysis (PCA)	HR measurement.
[26]	X-iPPGNet; 3D Depthwise Separable Convolutions; Xception-inspired architecture	Pulse Rate (Heart Rate) estimation.
[24]	HRUNet; Bayesian Neural Network (BNN); STFT; Variational inference; CNN	HR measurement; Arrhythmia screening (Atrial Fibrillation)
[13]	TST-SFA (Time–Space Transformer); Multiscale Retinex (MSR); ResNet-101	HR measurement
[14]	Systematic review: CNN, LSTM, SVM, KNN, Random Forest, BPNN	HR, HRV, Respiratory Rate, Drowsiness, and Fatigue monitoring
[23]	HSRD (Hierarchical Style-aware Representation Disentangling), ResNet-18, AdaIN	HR Estimation, HRV
[6]	CNN with Attention Mechanism (HRV-B, HRV-M, Dual-M), POS-based rPPG, OpenFace	Atrial Fibrillation (AF) Screening, HRV indices
[20]	CiSSA (Circulant Singular Spectrum Analysis); JBSS; IVA-G; CHROM	HR estimation
[8]	DD-ROI (Data-Driven ROI); Delaunay Triangulation; MediaPipe; CHROM; POS; PBV.	HR estimation
[27]	Fusion Video Vision Transformer (Fusion ViViT), Contrastive Learning (SSL)	Remote HR Estimation (Near-instant ~6 s)
[29]	ICA with Savitzky–Golay filter, Butterworth filter, nACF, and SAM	HR measurement from non-facial regions (forearm)
[22]	PhysNet; RTPPG (3D-CNN); KDE (Kernel Density Estimation); Image animation	HR estimation in NIR and Fitness
[7]	Adaptive-weight network (1D CNN), LSTM (3-layer), dlib landmark detector	HR Estimation, HRV
[21]	PhysFormer, Temporal difference transformer	HR Estimation, rPPG prediction
[25]	2D Variational Mode Decomposition (2D-VMD), AAPSD, Multimode Kurtosis, ICA	HR Estimation
[5]	Project_ICA (Signal projection+ICA), KLT tracker, Viola-Jones detector	HR Estimation
[1]	FastICA (Independent Component Analysis); FFT; Jacobi SVD; Fixed-point arithmetic	HR detection
[4]	rPPG, Butterworth band pass filter, Peak counting	HR (Pulse Rate) Estimation
[18]	Logistic Regression (LR), Support Vector Machine (SVM), Decision Trees (DT)	Emotional State Classification (Meltdown, Frustrated, Happy, etc.)
[17]	Viola-Jones, AdaBoost, Haar features, Cascade of classifiers	Body Pattern Detection (Headache, Happiness, Hunger, Fear, Recreation)

Table 3. Key concepts and key results.

Ref	Key Concepts	Key Results
[12]	Self-supervised Learning, Vision Transformer (ViT), PC-STMap (POS/CHROM with STMap)	MAE: 4.52 bpm (VIPL-HR), 0.40 bpm (PURE), 0.29 bpm (UBFC-rPPG).
[30]	Deep Learning; Vision Transformer (ViT); Spatio-temporal modeling; High-dimensional reconstruction	MAE: 0.41 bpm (UBFC-rPPG), 0.54 bpm (COHFACE), 4.12 bpm (VIPL-HR). Pearson $r \geq 0.82$.
[28]	Deep Learning (MTTS-CAN), Python 3.9 framework, face tracking (Kalman filter, MTCNN), GPU acceleration	MAE < 1 bpm for top methods; POS and CHROM outperformed baseline.
[16]	Deep Learning (CNN), FCN-8S segmentation, Multimodal imagery framework, Noise reduction (SCNAU)	HR accuracy: 99.83%; Precision: 90.37%; HR RMSE: 5.26 bpm, PCC: 0.923.
[19]	Deep Learning (ResNet for BP, BlazeFace for detection), intensity-based rPPG, cloud-based processing	HR MAE: 1.73 bpm (TokyoTech), 3.95 bpm (PURE). BP MAE: 6.7 mmHg (SBP)/9.6 mmHg (DBP).
[2]	Deep Learning (PFLD), Multimodal Integration (video rPPG + cushion BCG), Pose Constrained Kalman Filter	MAE 3.12 bpm lower than POS alone; Pearson correlation $r = 0.85$
[3]	Signal processing (EWT and PCA) to isolate blood volumetric pulse (BVP)	MAE: 0.98 bpm (PURE), 1.06 bpm (UBFC), 1.99 bpm (COHFACE), 1.81 bpm (ASIPL).
[26]	Deep Learning; CNN(3D-CNN); Color channel decoupling; Xception network	MMSE-HR: MAE = 4.10, RMSE = 5.32; UBFC-rPPG: MAE = 4.99; MAHNOB-HCI: MAE = 3.17.
[24]	Deep Learning; Bayesian posterior estimation; Uncertainty assessment	Health-HR-NSR: MAE = 1.758; Health-HR-AF: MAE = 5.412; Outperforms SOTA when uncertainty > 0.4 is excluded.
[13]	Deep Learning; Space-shared/specific features; Transformer; Affinity variation loss	VIPL-HR: MAE = 4.39; COHFACE: MAE = 1.31; BH-rPPG: MAE = 2.73.
[14]	Deep Learning, CNN, LSTM, Multimodal Integration (physiological + behavioral data)	Accuracy: Fatigue alerts 93.4% (mCNN); Drowsiness classification 91% (CNN-LSTM); HRV-based BPNN 88%.
[23]	Deep Learning, CNN(ResNet), Adversarial Learning, Domain Generalization, Representation Disentangling	MAE: 1.05 (PURE), 8.00 (VIPL-HR), 0.54 (UBFC), 8.66 (V4V)
[6]	Deep Learning, CNN, Attention Mechanism, Multi-task Learning, Motion Analysis	Sensitivity: 96.62%, Specificity: 90.61%, AUC: 0.96.
[8]	Supervised ROI selection; Facial mesh modeling; Dynamic skin segmentation	MAE: 0.55 bpm (PURE), 2.24 bpm (IIP-W), 6.19 bpm (IIP-F). Significant reduction in MAE vs. standard ROI
[27]	Deep Learning, Video Vision Transformer, Self-Supervised Learning, RGB-NIR Fusion, Multimodal Integration	RMSE: 14.86 bpm (VIPL-HR SSL); 16.94 bpm (MR-NIR-Car Transfer learning)
[29]	Non-facial iPPG, Signal processing (SG filter, SAM), robust to AC fluorescent light interference	Reduced errors by 83% for fluorescent lighting. MAE: 3.32 bpm under fluorescent tubes.
[22]	Deep Learning; Data Augmentation; Synthetic video generation; Near-infrared (NIR); Transfer Learning	IMVIA-NIR: MAE = 3.25 bpm; ECG-Fitness: MAE = 9.32 bpm; SNR improvement: −10.9 to 4.7 dB
[7]	Deep Learning, 1D CNN, LSTM, Multiple ROI Integration	MAHNOB-HCI: RMSE = 7.65 bpm, Standard Deviation (STD) = 7.55, Pearson Correlation (R) = 0.82.
[21]	Self-supervised Learning (SSL), Transfer Learning, Attention mechanism, Multichannel input	MAE: 4.97 (COHFACE), 1.71 (UBFC1), 1.62 (UBFC2). SSL improved performance on COHFACE
[25]	Spatial–Temporal Filtering, Variational Mode Decomposition, Statistical Mode Selection	COHFACE (Natural light): RMSE = 2.51 bpm, $r = 0.99$. Private dataset: RMSE = 0.80 bpm, $r = 0.99$
[5]	Blind Source Separation (ICA), Skin Reflection Model, Face Tracking, Color Normalization	Stationary MAD: 3.30 bpm; Computer Interaction RMSE: 7.10 bpm. Pale skin ($r = 0.98$) vs. Dark skin ($r = 0.50$).
[1]	FPGA Hardware Acceleration; Verilog HDL; ROI search; Real-time monitoring	ME: -0.76 ± 5.09 bpm (16 s window); Hardware computation time: 0.034 ms to 0.710 ms.
[4]	Remote PPG, RGB Channel Comparison, ROI (Forehead) extraction	Accuracy: Blue channel 89.09%, Red 79.22%, Green 76.82%.
[18]	Wearable Sensor (Empatica E4), Multimodal Integration (HR, HRV, EDA, TEMP, ACC)	Global accuracy 68%; Person-dependent accuracy up to 85%.
[17]	Machine Learning, Intelligent Agent, Mobile Application Integration, Boosting technique	Accuracy: Headache 77%, Happiness 75%, Hunger 82%, Fear 88%, Recreation 77%.

Table 4. Comparison of rPPG methods based on technical setup, population, validation, and other factors.

Ref	Technical Setup (Camera Type, Wavelength, Frame Rate (fps))	Signal Processing (ROI Strategy, Algorithm)	Dataset and Population (Dataset, Sample Size (<i>n</i>), Age Group, Skin Tone Diversity)	Experimental Conditions (Motion Protocol, Illumination Condition)	Validation and Metrics (Reference Device (Gold Standard (e.g., ECG, PPG)), Outcome Metrics (MAE, RMSE))	Significance and Limitations (Key Limitation, Clinical Relevance)
[6]	Logitech C920 webcam/CCD camera, RGB, 30/84 fps	IBI consistency, HRV indices; CNN-based Attention Network	657 participants (Largest AF database); Mean age $\approx$ 71.67 years; Taiwanese clinical sites	Talking, facial expressions, head movements; Ambient fluorescent lighting (200–400 lx)	Single-lead Sigknow EZYPRO ECG patch; Sensitivity 96.62%, Specificity 90.61%	Contact-free Atrial Fibrillation screening; Clinical relevance for telemedicine; Limitation: privacy and dark ambient lighting.
[20]	Logitech C930, C310, RGB, 30 fps	Cheek ROI (standardized rectangular); HRUNet (Bayesian Neural Network) with integrated Fourier transform	Health-HR ($n = 800$), UBFC, OBF, PURE, MMSE-HR, COHFACE, VIPL-HR; Chinese ethnicity; diverse skin colors in MMSE	Undisturbed, Motion-disturbed (nodding, speaking, moving), Light-disturbed (dark/dim)	Medical-grade oximeter; MAE: 1.73 (undisturbed), 7.96 (motion), 3.25 (light) bpm	Quantifies measurement uncertainty; Limitation: Uncertainty may fail to flag periodic noise (e.g., running frequency).
[28]	RGB-video camera, Visible light, Variable (25, 30, 61 fps)	Face detection (MTCNN, Kalman), ROI (Patch/Skin thresholding); POS, CHROM, PCA, SSR, ICA, GREEN, LGI, PBV	PURE, LGI, UBFC, MAHNOB, COHFACE; $n = 6$ to 164; Age 19–40 (MAHNOB); Skin tone not in source	Resting, talking, head rotation, translation; Controlled and natural light	ECG, BVP; MAE, RMSE, PCC	Proposes open Python framework (pyVHR); Lack of standardized pre/post processing and reproducible evaluation.
[30]	Webcam or phone camera, RGB, 25 fps	Facial frame differences; TranPulse (Two-stage Video Swin Transformer)	UBFC-rPPG, COHFACE, VIPL-HR ($n = 107$), PURE; Imbalanced population distributions	Head rotation, talking, expressions; Studio, bright, and dim lighting	Realistic heart waveforms (PPG/ECG); MAE: 4.69 (VIPL-HR), RMSE: 7.53 (VIPL-HR)	Disentangles multiphysiological interference (e.g., respiration); Limitation: slightly worse performance in dim light.
[26]	RGB cameras, 25–61 fps	Face segmentation and cropping; X-iPPGNet: 3D Depthwise Separable Convolutions	BP4D+ ($n = 140$), MMSE-HR ($n = 40$), MAHNOB-HCI ($n = 27$); Highly diverse (Black, White, Asian, Hispanic)	Spontaneous facial expressions, significant head motions, occlusions	Contact sensors (BVP/ECG); MAE: 4.10 (MMSE), 4.99 (UBFC), 3.17 (MAHNOB) bpm	End-to-end PR estimation from 2s windows; Limitation: No BVP signal recovery prevents pulse wave feature analysis.
[29]	Eyepiece industrial microscope, RGB, 25 fps	Non-facial skin (forearm); Savitzky–Golay filter, ICA, FFT spectrum accumulation mechanism (SAM)	SF-VS dataset, $n = 83$; Age 18–74; Fitzpatrick types II to V	Minimal movement; Ring LED, LED downlight, and ceiling fluorescent tubes	Arduino-based pulse sensor; MAE: 0.73 to 3.44 bpm	Addressed AC light interference in non-facial regions; Limitations: lacks Fitzpatrick Type-I; misses some clinical requirements.
[8]	Microsoft Lifecam Studio/Logitech C920, RGB, 20–61 fps	Delaunay Triangulation (898 triangular ROIs); Data-driven ROI (DD-ROI) selection	PURE, IIP-F, IIP-W, COHFACE, UBFC, MAHNOB-HCI; Asians and Caucasians	Resting, talking, facial rotation; Varying light intensity	Contact PPG/ECG; MAE: 1.70 (PURE), 2.24 (IIP-W)	Systematic analysis of facial regions; Recommendation of forehead/cheek ROIs; Limitation: Occlusion (hats/hair).
[23]	Commodity cameras, RGB, 30 fps	STMap, Hierarchical Style-aware Representation Disentangling (HSRD) using ResNet-18	VIPL-HR, V4V, PURE, BUAA, UBFC-rPPG; Diverse skin color and gender	Varying illumination, different movement levels, complex backgrounds	ECG/PPG (CMS50E); MAE: 4.31 (UBFC), RMSE: 6.30 (UBFC)	Addresses domain shift and instance-specific variation; Limitation: imprecise domain categorization in implicit DG
[3]	Webcam (Logitech B525), Smartphone (Nord2), RGB, 20–30 fps	Sub-ROIs (25 $\times$ 25); Homomorphic filtering, EWT (Empirical Wavelet Transform), and PCA	PURE, UBFC, COHFACE, ASIPL ($n = 45$); High diversity including darker skin tones (India)	Natural and artificial light; rigid/nonrigid head motion and facial expressions	Finger pulse oximeter (CMS-60C); MAE: 0.98 (PURE), 1.06 (UBFC), 1.99 (COHFACE), 1.81 (ASIPL) bpm	Substantial tolerance to varying skin tones and motion; Limitation: rapid facial expression changes.
[16]	Logitech C920 HD Pro, RGB, 30 fps	ROI (face, nose); Cyclical algorithm (PCA + FFT), Cascade residual CNN-FPNR, ICNet	UBFC-RPPG, $n = 40$; Age 18–35; Caucasian, Asian, Hispanic (von Luschan chart)	Indoor sitting, spontaneous movements; Sunlight and fluorescent ceiling lamps	CMS50E pulse oximeter (PPG); MAE, RMSE, MSE, PCC	Predicts multiple vitals (HR, RR, SpO ₂); Accuracy increases when adding trends to models.
[19]	Smartphone camera, Visible (RGB), 30 fps	MediaPipe Face Mesh (478 landmarks), ROIs (Forehead, cheek, nose); ICA, ResNet	TokyoTech ($n = 9$), PURE ($n = 10$), Video-HR ($n = 15$), Video-BP ($n = 49$); Age 10–80 years	Steady, talking, rotation, daily living environment; Ambient lights	Finger PPG, Pulse oximeter, Andesfit BP Monitor; MAE	Validated in daily living environments; Commercialized as ‘Veyetals’ mobile app (initial release version, 2022).
[25]	Webcam (QHMP), Green channel, 6–20 fps	ROI (Forehead, cheeks); 2D-VMD (Variational Mode Decomposition) with AAPSD	Private ($n = 25$, 28 ± 2 yrs) and COHFACE ($n = 40$, 35 ± 11 yrs); Variable skin tones	Sitting, no large movements; Studio (Halogen) and natural light	Contact PPG (Nonin SenSmart); ME, RMSE, SD, Correlation (r)	2D-VMD reduces error significantly compared to ICA; Dlib face detection fails during head rotations.

Table 4. Cont.

Ref	Technical Setup (Camera Type, Wavelength, Frame Rate (fps))	Signal Processing (ROI Strategy, Algorithm)	Dataset and Population (Dataset, Sample Size (N), Age Group, Skin Tone Diversity)	Experimental Conditions (Motion Protocol, Illumination Condition)	Validation and Metrics (Reference Device (Gold Standard (ej. ECG, PPG)), Outcome Metrics (MAE, RMSE))	Significance and Limitations (Key Limitation, Clinical Relevance)
[22]	RGB and Near-infrared (NIR) cameras; 30 fps	Face tracking (BlazeFace) and skin detection; Physnet and RT-PPG (3D-CNN)	MERL-RICE, TokyoTech-NIR, ECG-Fitness, IMVIA-NIR ($n = 10$, varying ethnicities)	Fitness activities (rowing, running, biking), low-light/nighttime (NIR)	Pulse oximeter, finger PPG sensor; MAE (lowered to 1.63 bpm with DA on MERL)	Synthetic video generation for data augmentation; Addresses scarcity of NIR and movement-heavy datasets
[5]	Logitech C270i, RGB, 30 fps	Face tracking/Skin detection; Project_ICA (Simplified skin reflection model)	28 participants (18 pale, 10 dark skin); 112 videos	Stationary, computer interaction, swinging heads, exercise recovery; External daylight	Finger pulse oximeter (YUWELL YX303); MAD (3.30 stationary), RMSE (7.21 stationary)	Robust against head movement; Key limitation: Obtaining reliable measurements from dark-skinned subjects.
[21]	High-resolution camera, RGB, 30 fps	Facial regions (cheek, forehead, jaw); Physformer (Video Transformer) with SSL	COHFACE, UBFC1, UBFC2, PURE ($n = 10$); Various skin tones	Talking, reading, exercising; Different lighting and camera angles	Contactless imaging PPG; MAE: 4.97 (COHFACE), RMSE: 5.71 (COHFACE)	Reduces reliance on labeled data through SSL; Limitation: performance on Pearson's Correlation was not improved
[1]	TRDB-D5M camera, RGB, 16 fps	Skin detection rules; ICA implemented on FPGA	12 subjects, 142 data points, Age 21–28, Male/Female	Resting, mitigated motion artifacts; Ambient lighting	Omron HEM-6111; Accuracy (ME $\pm$ 1.96SD) of -0.76 ± 5.09 bpm (16 s)	Fast FPGA hardware accelerator for real-time edge processing; Limitations: Fixed-point precision inferiority.
[7]	Consumer-level camera, RGB (Green channel), 30 fps	Multiple cheek ROIs (annulus/circular); Adaptive-Weight Network with LSTM	MAHNOB-HCI, 27 subjects, Age 19–40; Slight facial expressions	Emotional stimulation scenarios; Slight facial expressions, small body movements	ECG; RMSE 7.65 bpm, Pearson's R 0.65	Dynamically selected ROI weights; Limitation: ROI loss during head turning.
[4]	Mobile phone camera, RGB, 30.35 fps	Manual static forehead ROI; Color Intensity Pulse Count (Butterworth band-pass filter)	DMIMS database, $n = 20$; Women aged 20–35; Skin tone diversity not specified	No motion protocol (subjects held hands up), Controlled environment	MP20 monitor with ixTrend software (Version not available); Accuracy (Blue channel: 89.09%, Red: 79.22%, Green: 76.82%)	Low accuracy due to lack of SNR improvement; Demonstrates feasibility via mobile phone without specialized hardware.

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only

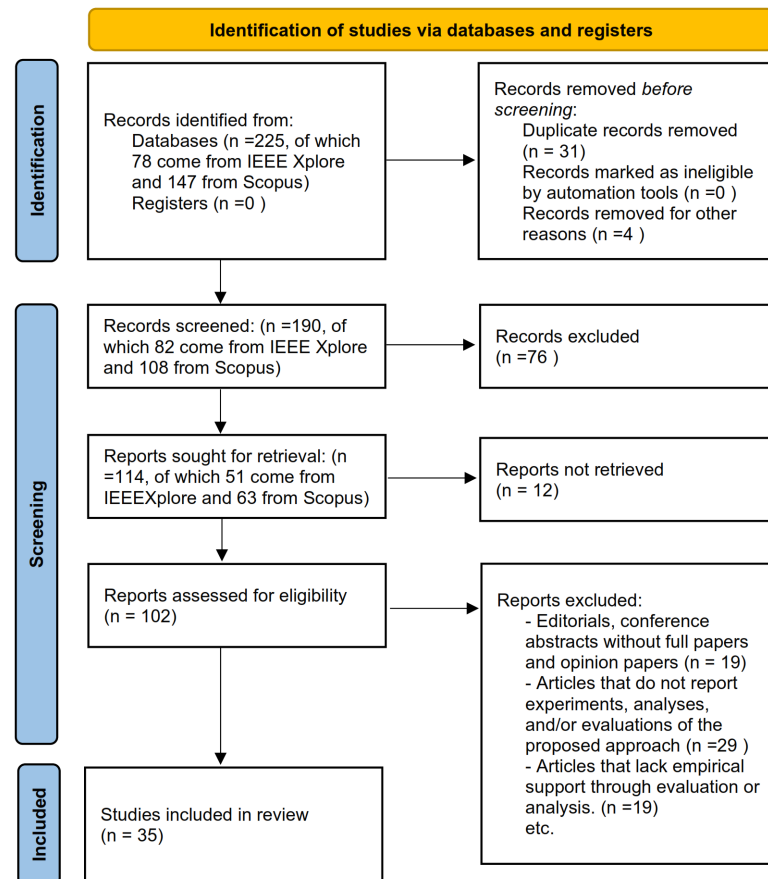


Figure 1. Prisma 2020 flow diagram.

3. Fundamentals of Remote Photoplethysmography (rPPG)

3.1. Physiological and Optical Principles

The technique known as rPPG is based on the ability to detect subtle variations in the intensity of light reflected from human skin, which are caused by changes in capillary blood volume throughout the cardiac cycle [23]. This phenomenon, referred to as the photoplethysmography (PPG) signal, relies on the optical properties of hemoglobin, which exhibits higher light absorption compared to surrounding tissues [2]. By utilizing conventional RGB cameras, it is possible capture these variations across the color channels (red, green, and blue). Recent research has highlighted that the analysis of specific channels can provide higher accuracy depending on the environment and capture technology (such as mobile phone cameras); for instance, the blue channel has been identified as yielding highly precise results (up to 89.09%) when counting color intensity peaks for HR estimation in uncontrolled settings [4].

3.2. Evolution of Algorithms: CHROM, POS, ICA

rPPG signal processing has evolved from traditional statistical approaches toward more robust signal processing algorithms. Among the most widely used methods for separating the rPPG signal from noise and illumination variations are ICA, CHROM, and POS, as well as models based on the PBV [20,28]. Evaluating the technical strengths and limitations of these mathematical frameworks is essential when considering their deployment in vulnerable pediatric populations, such as children with CP, whose involuntary movements and spasticity present severe challenges for signal stability and subsequent affective monitoring:

- ICA: This method utilizes blind source separation (BSS) to decompose temporal RGB color mixtures into independent non-Gaussian components. Although revolutionary for isolating the cardiac pulse from non-periodic noise without prior training, its accuracy significantly degrades under rapid motion. Studies show that ICA requires higher processing time due to its iterative nature and is prone to “component switching,” where the pulse signal moves between different independent channels [28,29]. This limitation is critical in children with CP; sudden spastic movements or dystonic posturing can cause continuous component switching, making tracking stable HRV features for stress or pain classification highly unreliable.
- CHROM: By assuming a standardized skin color to define a projection direction, this algorithm eliminates specular reflection and mitigates ambient light variations. Technically, it offers a superior balance between complexity and accuracy; it is computationally lighter than ICA and shows lower MAE in scenarios with stable illumination. However, its reliance on fixed skin-tone ratios limits its performance across diverse phenotypes [20,28]. For pediatric neurorehabilitation or home-care settings, ambient light often fluctuates as the child moves or changes posture, meaning that while CHROM reduces minor motion artifacts, it may fail to isolate the micro-vascular pulse during intense affective distress episodes where the child exhibits heightened motor activity.
- PBV: This model leverages the specific “signature” or vector of blood volume changes across the RGB spectrum. By using a pre-defined PBV, it achieves higher robustness against motion artifacts than CHROM. In comparative tests, PBV maintains a stable Signal-to-Noise Ratio SNR even when the subject is talking, although its precision depends heavily on the correct calibration of the camera’s spectral sensitivity [12,28]. This mathematical resilience represents an advantage for children with CP who present uncoordinated facial expressions or vocalizations; however, the reliance on precise spectral calibration limits its clinical feasibility across different standard webcams or video settings used in ecological or school environments.
- POS: Representing the current state-of-the-art in mathematical methods, POS defines a projection plane orthogonal to the skin tone in a temporally normalized RGB space. According to recent benchmarks, POS consistently achieves the lowest RMSE in datasets with large-head rotations and exercise (e.g., $MAE \approx 2\text{--}5$ bpm in moderate motion). It outperforms CHROM and ICA by effectively isolating the pulse from intensity variations caused by distance changes to the sensor [2,28]. Consequently, POS emerges as the most clinically feasible algorithmic candidate among traditional methods for children with CP, as it is capable of handling the frequent distance and orientation changes relative to the camera caused by trunk instability or wheel-chair adjustments, while maintaining the HR quality required for objective emotion monitoring.

3.3. Recent Advances with Deep Learning

The integration of Artificial Intelligence has transformed the field, enabling the management of complex and non-linear data. Through the training of deep learning models for HR estimation, it is possible to effectively expand and augment datasets [22]. However, adapting these data-driven architectures to neurodevelopmental conditions such as CP requires a careful examination of how specific deep models handle severe motion, non-standard facial geometries, and atypical behavioral responses linked to emotional distress.

- CNN: Primarily utilized for extracting spatial-temporal features directly from video frames. Architectures like X-IPPGNet leverage depthwise separable convolutions to reduce the number of parameters, achieving a high precision with a RMSE of 6.26 bpm and a MAE of 4.99 bpm on the UBFC-rPPG dataset [26]. This balance between

accuracy and efficiency enables precise ROI segmentation even in less-constrained scenarios [6,26]. For children with CP, the lightweight and robust nature of these spatial-temporal CNNs is crucial, as they can maintain ROI tracking even when involuntary facial movements or hypertonia alter standard facial landmarks, ensuring consistent pixel extraction for pulse analysis.

- **LSTM:** These recurrent networks are essential for modeling the sequential nature of cardiac signals, assisting in the prediction of HR over time. For instance, recent architectures utilize an LSTM model in a multi-region framework to effectively map optimized remote photoplethysmography signals from multiple facial ROIs into precise HR estimations. This approach significantly mitigates instabilities caused by motion artifacts and illumination variations by capturing long-term temporal dependencies in the pulse signal, balancing accuracy against the inherent computational latency of recurrent structures [7]. In affective monitoring for pediatric CP, LSTM networks provide a valuable framework to differentiate between sudden physiological spikes caused by a physical spasm and the prolonged autonomic changes associated with persistent pain or emotional distress.
- **Transformers:** Their global attention mechanism allows for capturing long-term dependencies within the video signal. While they require larger datasets for training, they can outperform CNNs in pulse wave reconstruction by focusing on relevant skin patches, thus improving the SNR during significant head rotations [12,27]. This self-attention capability is highly advantageous for children with CP, who frequently experience sudden head drops, lateral rotations, or partial facial occlusions due to wheelchair restraints or involuntary posturing; the model can dynamically shift its attention to unoccluded skin regions to prevent signal dropout.
- **Self-supervised learning:** Frameworks like rPPG-MAE address the scarcity of labeled clinical data by using pre-training on unlabeled videos. Adopting this approach enhances system generalization and narrows the “domain gap,” enabling models with robust performance across different lighting conditions and camera sensors without requiring synchronized ECG ground-truth during initial training [12,21,27]. This learning paradigm directly addresses a major bottleneck in pediatric clinical research, where collecting synchronized gold-standard physiological data (e.g., ECG or contact sensors) from highly agitated or uncooperative children with severe CP is often clinically and ethically unfeasible.

3.4. Technical Challenges

Despite advancements in the implementation of rPPG, several critical challenges remain to be addressed. Key hurdles include:

- **Motion artifacts:** Both voluntary and involuntary movements introduce noise that can mask the pulse signal. Hardware accelerators have been proposed to mitigate these effects in real-time [1,2,12].
- **Illumination changes:** Variations in ambient light alter skin reflectance, complicating the consistent detection of intensity peaks [13,31].
- **Skin tone variability:** Melanin levels affect light absorption, which can introduce biases in the accuracy of traditional algorithms [12].
- **Limited datasets:** There is a notable lack of extensive, public databases that include diverse populations, specifically children with CP, to validate the full reliability of these systems in real world environments and ensure their integration into proactive monitoring systems [14].

4. Applications of rPPG in Physiological Monitoring

rPPG technology has proven to be a versatile tool for non-invasive health monitoring, enabling the extraction of critical physiological parameters from subtle skin color variations captured by conventional cameras. In pediatric populations with CP, deploying these systems requires a specialized pipeline designed to transform raw video data into reliable vital signs despite severe motor challenges. The implementation of these systems follows a specialized pipeline designed to transform raw video data into reliable vital signs. This process begins with the identification and tracking of specific facial ROI, where recent advancements use triangulation and adaptive tracking for signal integrity during movement [8,32]. Once the ROI is established, the signal is extracted by analyzing intensity variations across color channels; while the green channel is traditionally preferred [4], modern implementations often incorporate blue channel analysis [1] or multi-spectral fusion (RGB-NIR) to enhance accuracy in uncontrolled lighting conditions [27,31]. The core implementation utilizes signal processing architectures—ranging from ICA [5] and non-parametric models [20] to deep learning architectures like X-iPPGNet [26]—to isolate the blood volume pulse from environmental noise and motion artifacts [9,13]. However, the precision of these implementations can be compromised by social media beauty filters that alter facial features and skin appearance, impacting both recognition and the resulting physiological metrics [33]. The following subsections describe how these implemented frameworks are applied to specific vital signs.

4.1. Heart Rate (HR) and Heart Rate Variability (HRV)

HR estimation represents the most consolidated application of rPPG. The use of advanced algorithms and synthetic data augmentation has significantly enhanced the precision of these systems, even in challenging scenarios [5,25]. According to the recent study [22], the implementation of robust rPPG frameworks achieves a MAE of 1.3 beats per minute (bpm). While these figures demonstrate promising accuracy for heart rate estimation, achieving clinical-grade reliability in complex scenarios, such as monitoring children with CP, remains an open challenge due to potential motion artifacts, atypical postures, and involuntary movements.

Regarding HRV, frameworks such as pyVHR facilitate the estimation of these fluctuations from videos, which is fundamental for safety and well-being applications [15]. The analysis of HRV through non-intrusive instrumentation has been identified as a promising indicator for the continuous monitoring of fatigue and drowsiness in drivers, enabling the creation of proactive alert systems [34].

4.2. Respiratory Rate and Blood Oxygenation

Beyond heart rate, rPPG enables the prediction of other essential vital signs, such as breathing rate (BR) and arterial oxygen saturation (SpO_2), using video captured under ambient light [16,19]. The implementation of CNN and Long Short-Term Memory (LSTM) models has enhanced the robustness of these estimations against environmental distortions and involuntary movements [6,7]. Specifically, it has been proposed that optimized remote photoplethysmography (iPPG) signals possess the potential to estimate complex health indicators within lightweight systems suitable for embedded devices [7].

4.3. Multimodal Estimation (Blood Pressure, SpO_2 , Multiple Vital Signs)

Current trends are moving toward multimodal fusion to overcome the limitations of individual sensors. A technical development in this direction involves the integration of rPPG with BCG, which detects body movements induced by the cardiac cycle [2]. This combination exploits the differential sensitivities of both techniques: while rPPG is

susceptible to facial movements, BCG is sensitive to body movements, allowing for robust HR estimation even during high-motion states. For children with CP, this multi-sensor framework offers an excellent compromise: when the face is temporarily occluded or turned away during a spasm, the mechanical BCG signal captured from the wheelchair or seat sensors can bridge the data gap.

The Oxygen Saturation (SpO_2) estimation is based on measuring light absorption across different wavelengths. Since oxygenated and deoxygenated hemoglobin have distinct absorption spectra, rPPG can calculate saturation levels by analyzing the ratio of the Direct Current (DC) and Alternating Current (AC) components of signals captured in specific spectral channels, such as red and near-infrared [19], requiring heuristic tuning of certain parameters. In patients with CP, tracking SpO_2 alongside HR provides a more holistic view of affective and physical distress, distinguishing pure emotional arousal from respiratory depression related to motor impairment.

The Respiration Rate (RR) can be extracted from both rPPG and BCG. In rPPG, respiration modulates the cardiac signal through variations in amplitude and frequency (respiratory sinus arrhythmia). In BCG, the captured signal is a direct combination of cardiac and respiratory mechanical activities, which can be isolated using specialized filtering techniques [2,22].

While not measured directly by the camera, Blood Pressure (BP) can be estimated indirectly using Pulse Transit Time (PTT). By combining rPPG (marking the arrival of the pulse at a peripheral site, like the face) with BCG (marking the start of cardiac ejection), PTT can be calculated and correlated with systolic and diastolic blood pressure through physiological models [2]. In clinical and home-care settings for children with severe disabilities, tracking non-contact continuous BP changes provides an additional safety layer for assessing chronic stress and nociceptive (pain) activation without triggering the tactile discomfort associated with traditional inflatable cuffs.

4.4. Validation Against Reference Methods (ECG, Pulse Oximetry)

To establish the clinical and technical reliability of rPPG systems, the literature emphasizes a multi-level validation approach against gold-standard physiological measurements. This validation is critical to guaranteeing that non-contact signals remain accurate under the physical constraints of the target population:

- **Reference software:** The use of specialized tools such as ixTrend allows for the comparison of pulse rates extracted from video with actual heart rates recorded simultaneously [4].
- **Medical devices:** Results are typically validated against commercial oximeters (such as the Omron HEM-6111) for the accuracy of the obtained bpm values [1].
- **ECG signals:** ECG remains the primary gold standard for validating the temporal precision of rPPG. Its role is critical not only for basic HR estimation but for calculating HRV. In children with CP, ECG provides the necessary timing accuracy to guarantee that rPPG reliably captures physiological shifts associated with emotional dysregulation or pain [2].
- **Statistical metrics:** Validation is supplemented by MAE analysis and Mean Absolute Percentage Error (MAPE), frequently visualized through Bland–Altman plots [32].

5. rPPG and Emotion Monitoring. Particularization to Children with Cerebral Palsy

Monitoring the emotional state of the pediatric population with CP presents unique challenges that rPPG aims to address through non-intrusive observation.

5.1. Specific Challenges: Spasticity, Limited Communication, and Motor Variability

Wearable devices have been successfully utilized to monitor physiological signals in school-aged children with the objective of identifying emotions [18]. However, children with CP often exhibit spasticity and involuntary movements that complicate the use of contact sensors (such as ECG chest straps), as these can cause discomfort or become detached due to motor variability. Furthermore, the limited communication abilities of many patients prevent the verbal expression of pain or distress states. rPPG offers a robust alternative by capturing physiological signals from a distance, thereby avoiding physical contact and reducing patient anxiety [18].

This non-invasive alternative becomes even more critical when considering that atypical facial geometry and unpredictable motor patterns often impose severe constraints on conventional machine learning frameworks. Indeed, automated affect classification frameworks applied to infants at risk of neurodevelopmental delays have demonstrated that deep neural networks suffer sharp performance drops when confronted with the geometric variability inherent to spontaneous, unconstrained movement [35]. Crucially, evidence suggests that lowering the problem's dimensionality—by relying on essential facial concavity features rather than dense landmark meshes—can yield much higher statistical stability against motor distortions. Transposing these insights to cerebral palsy monitoring suggests that instead of forcing massive, normatively trained computer vision models to adapt to clinical cohorts, the key to mitigating spasticity-induced motion artifacts lies in isolating local, robust morphological descriptors capable of tolerating transient asymmetries during motor surges or dystonic posturing.

5.2. Emotion and Stress Recognition from HR/HRV

The literature suggests a direct correlation between the activity of the Autonomic Nervous System (ANS) and the parameters of HR and HRV. The monitoring process is based on the principle that emotional states—such as stress, pain, or frustration—trigger a physiological response characterized by a shift in the sympathovagal balance. According to the systematic review by [14], HR analysis is fundamental for detecting alertness levels and fatigue states, providing a robust baseline for physiological monitoring. In children with CP, where verbal communication is often limited, these indicators provide an objective “voice” for their internal state.

The implementation of emotional monitoring through these indicators follows a structured physiological mapping:

HR as an alertness indicator: An acute increase in HR, detected via rPPG, serves as a primary marker for high-arousal states. In clinical and home settings, sudden spikes in pulse rate can be correlated with states of physical discomfort, acute pain, or environmental overstimulation [18,30].

HRV for emotional regulation: HRV represents the millisecond-to-millisecond variations between heartbeats and is a more sensitive marker for cognitive load and emotional regulation. A decrease in HRV (specifically in time-domain metrics like RMSSD) indicates a withdrawal of parasympathetic (vagal) tone, which is strongly associated with states of stress, anxiety, or emotional dysregulation in pediatric populations [15,18].

Correlation with behavioral patterns: To enhance precision, these physiological shifts are mapped against non-verbal behaviors. For instance, the use of specialized frameworks for children [17] allows for the identification of physiological shifts associated with emotional dysregulation, even when clear behavioral signs are absent. This is further supported by the integration of rPPG with motion analysis to differentiate between physical exertion and genuine distress [2,24].

By utilizing frameworks like pyVHR [15] to extract these fluctuations from video, rPPG systems can identify these physiological signatures, transforming raw pulse data into actionable emotional insights.

5.3. Integration with Complementary Biomarkers and Limitations

To overcome reduced facial expressivity and motor variability in children with CP, current research points toward multimodal monitoring systems.

An important development in this field is the work of [17] regarding the use of computer vision for body expression detection. The study demonstrates that it is possible to identify, with high precision, movement patterns associated with states of well-being and necessity, such as hunger (82%), fear (88%), and pain (77% for headaches). By integrating these behavioral detections with HR variations obtained via rPPG, a more holistic monitoring framework is established. This allows caregivers to proactively interpret the child's needs through digital interfaces, compensating for the limitations in verbal communication.

Likewise, the inclusion of oculometry as a complementary biomarker offers a direct window into emotional and cognitive states non-invasive, remote sensing technology. According to the research by [10], the use of remote video-based eye trackers allows for the precise analysis of pupil diameter, saccadic movements, and fixations without requiring physical contact with the user. In this study, the emotional states are identified by monitoring these ocular responses while participants interact with a standardized set of visual stimuli designed to trigger specific levels of pleasure and physiological activation. This methodology demonstrates that complex emotions, such as fear and sadness, can be differentiated by capturing ocular data from a distance, much like facial imaging. Since ocular control is often one of the best-preserved motor functions in patients with severe motor disabilities, its integration with rPPG enables the correlation of the body's physical response with the emotional state detected through gaze. This synergy between remote oculometric data and pulse signals is fundamental for the development of future assistive technologies designed with the aim of improving quality of life and personal expression in both clinical and domestic settings.

However, critical limitations persist. The lack of sufficient validation in real-world environments and the need for algorithms capable of processing signals under extreme motion conditions remain significant barriers [25]. As noted in the review by [14], the full reliability of these technologies for proactive emotion detection still requires more extensive clinical research and specific databases tailored to the population with motor disabilities.

6. Discussion: Critical Synthesis, Identified Gap and Future Perspectives

The current literature review reveals a clear dichotomy in the state of the technology: while rPPG has reached significant maturity for physiological monitoring (HR/HRV), its application for emotion detection in populations with neurodevelopmental disorders remains in an incipient stage.

The analysis of the studies summarized in (Table 2) reveals that research has mainly focused on HR estimation, with a growing but still limited interest in multimodal estimation. Regarding the technical approach, the most frequently implemented algorithms for signal extraction and noise reduction are POS, CHROM, and ICA, alongside modern deep learning architectures such as CNN and LSTM. Additionally, for ROI detection and tracking, the literature shows a consistent reliance on classical but robust methods like Viola-Jones, AdaBoost, and Cascade of Classifiers, which remain relevant for real-time applications despite the rise of more complex neural networks.

6.1. Maturity of rPPG vs. Immaturity for Emotion Detection

The accuracy achieved in the estimation of HR, particularly through optimizations such as blue channel analysis (89.09% accuracy, while the red channel 79.22% and the green channel 76.82%) [4] demonstrates that rPPG is already a viable alternative to contact sensors for basic telemetry. However, translating these signals into emotional states remains a complex challenge. Although pulse variability serves as an indicator of emotional intensity (arousal), its specificity is limited. In clinical contexts, such as with children with CP, interpreting these data is even more difficult; physiological responses may be altered by external factors such as medication (which can mitigate heart rate variability) or the quality of the captured signal due to a low signal-to-noise ratio if illumination corrections or motion stabilization are not applied.

6.2. Potential of Multimodal Integration

The primary opportunity for advancing emotion recognition lies in multimodal integration, specifically through non-invasive techniques that ensure user comfort. Although traditional ECG or contact PPG offer high precision, they require skin-attached sensors that can be intrusive or impractical for children with motor disabilities. Instead, the synergy between rPPG with BCG [2] provides a quasi-contactless alternative. In this framework, BCG captures the mechanical vibrations of the body caused by heart contractions; although it is also susceptible to motion, its fusion with rPPG through pose-constrained filters allows the system to cross-reference signals and mitigate noise that would otherwise invalidate a single-source measurement. To further refine the emotional profile, this physiological data is triangulated with motor behavior. This includes body expression detection to identify physical agitation [17] and remote eye-tracking, which utilizes video-based analysis of pupil diameter and fixations as direct indicators of cognitive and emotional load [10].

6.3. Technical Robustness: Algorithms Resistant to Motion

The most critical technical “gap” for the pediatric population with CP is robustness against spasticity. Conventional algorithms typically fail when faced with sudden involuntary movements. However, the development of FPGA-based hardware accelerators [1] and model training through masked autoencoders in self-supervised learning [12] pave the way for real-time processing capable of dynamically tracking ROIs, thereby isolating the pulse signal even in scenarios of high motor instability.

Regarding the performance metrics detailed in (Table 3), the evidence confirms that the POS algorithm consistently yields the most robust results among non-contact methods, particularly when handling flickering light and moderate movement; while techniques like CHROM and ICA show competence in static conditions, POS maintains higher signal-to-noise ratios in more dynamic scenarios. However, the data also reveals a significant trend: the high precision reported across these studies (often exceeding 90%) is predominantly tied to datasets of healthy adults. This highlights a critical lack of evidence regarding the performance of these algorithms when faced with the involuntary motor patterns and physiological specificities of the pediatric CP population.

6.4. Methodological and Ethical Limitations

Despite these advancements, there is a notable scarcity of clinical studies specifically targeting children with CP. Most algorithms are validated on healthy adults or within controlled laboratory settings. To achieve ecological validity, it is imperative conduct validations in real environments. Ethically, the use of rPPG and computer vision in this population requires sensitive management of privacy and consent, ensuring that the technology reduces caregiver burden rather than introducing new surveillance concerns.

6.5. Expected Impact and Future Perspectives

Looking ahead, the goal is to transition from reactive to objective and proactive monitoring. The combination of rPPG, eye-tracking, and wearable sensors will transform communication for children with severe verbal limitations. The expected impact is threefold.

As future work, we propose the development of a multimodal fusion architecture that transcends the limitations of isolated monitoring identified in this review. The technical roadmap will focus on signal triangulation: physiological data obtained through the POS algorithm, selected for its superior robustness against lighting changes and motion compared to basic chromatic models, will be correlated with a contact sensor (wearable ECG) to establish a ground-truth and validate algorithmic precision against motion artifacts. The methodology will follow a progressive scaling: first, the system's stability will be validated in controlled environments with healthy volunteers; subsequently, it will be implemented in real-world settings with children with CP. Unlike the studies analyzed in (Table 3), which primarily focus on adult populations or laboratory conditions, our proposal directly addresses motor variability within the child's ecological environment. This approach will enable a transition from reactive monitoring to a proactive system capable of interpreting states of pain or well-being, positively impacting patient autonomy and significantly reducing the supervision burden on caregivers.

6.6. Translational Roadmap for rPPG Technology in Children with Cerebral Palsy

Bridging the gap between academic research in non-contact physiological sensing and its practical deployment for children with CP requires a structured translational roadmap (Figure 2). Accordingly, this review proposes a multi-phase framework (T1 to T4) designed to overcome the critical algorithmic, technical, and clinical bottlenecks previously identified.

- **Phase T1: Laboratory Validation and Algorithmic Robustness.**
The foundation of the roadmap focuses on mitigating the severe motion artifacts induced by spasticity and involuntary movements characteristic of pediatric CP; while classical mathematical frameworks like POS provide a baseline for handling head rotations under controlled constraints [2,28], deep learning architectures offer the necessary flexibility for clinical translation. Lightweight spatiotemporal convolutional neural networks (CNNs) represent a crucial step forward, allowing continuous tracking of regions of interest (ROIs) even when severe hypertonia distorts standard facial geometry [6,26]. To finalize this laboratory phase, future developments must prioritize recurrent networks capable of modeling long-term temporal dependencies to stabilize the pulse signal against sudden shifts [7], alongside the optimization of alternative chromatic channels that have already demonstrated high accuracy in unconstrained preliminary tests [4].
- **Phase T2: Clinical Testing and Addressing the Data Gap.**
The second phase marks the transition to controlled clinical evaluations, a step currently hindered by a critical methodological void: the absolute scarcity of public, labeled rPPG datasets involving children with neurodevelopmental disorders [14]. Forcing algorithms trained on neurotypical adults to interpret atypical infant or pediatric motor patterns causes severe performance degradation due to the extreme geometric variability of spontaneous motion [35]. To break this bottleneck without introducing highly intrusive contact-based reference sensors (like ECG) that stress agitated children, T2 translation must leverage self-supervised learning (SSL). Implementing frameworks such as masked autoencoders allows models to pre-train on massive unlabeled video data, effectively bridging the domain gap and enhancing generalization in actual clinical environments [12,21,27].

- **Phase T3: Multimodal Integration and Clinical Utility.**
Phase T3 shifts the focus toward multi-sensor systems capable of providing comprehensive clinical utility. Objective emotional and physiological monitoring in children with severe communication impairments cannot rely on a single optical modality. During sudden spasms or posture adjustments where the child's face is completely occluded, the rPPG signal must be dynamically supported by mechanical variables. Fusing rPPG with ballistocardiography (BCG) via sensors integrated directly into the wheelchair's seating or support structure ensures continuous, uninterrupted monitoring of vital signs like heart rate and indirect blood pressure [2]. This physiological telemetry must be cross-validated with robust local morphological descriptors that automatically identify gross motor expressions of pain or distress [17].
- **Phase T4: Real-World Community and Home Implementation.**
The ultimate goal of this translational roadmap is the permanent deployment of non-invasive monitoring in the daily lives of patients (homes and schools). To achieve true ecological validity, the autonomic markers (HR/HRV) extracted remotely via rPPG toolkits [15] must be translated into real-time, interpretable alerts for caregivers and families. This final phase will provide a reliable, objective "voice" to children with cerebral palsy, transforming a laboratory optical technique into an accessible assistive technology that directly improves their quality of life.

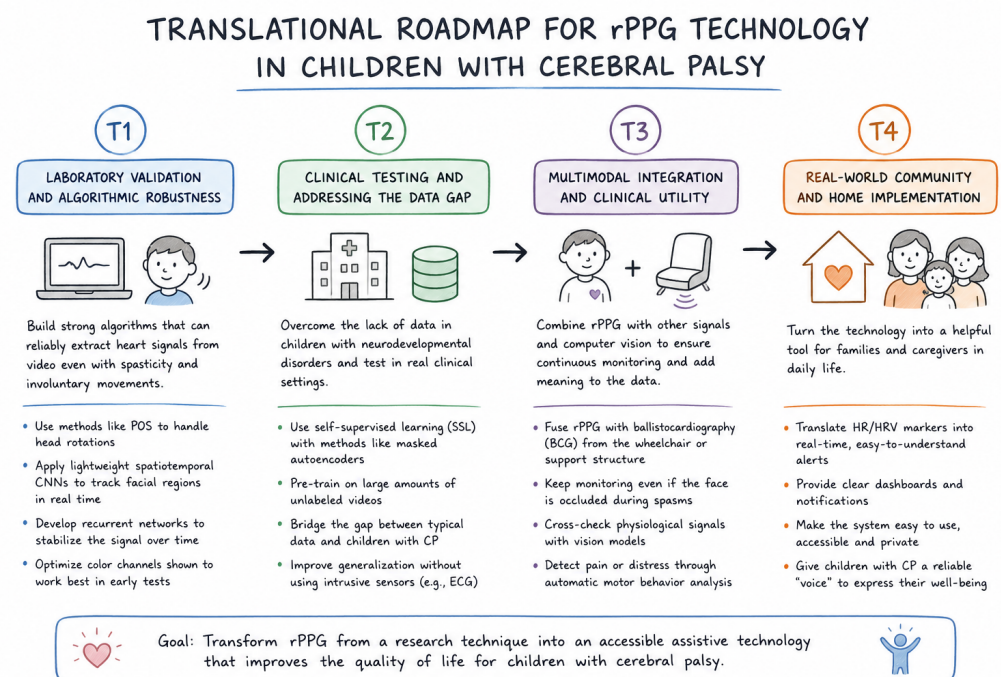


Figure 2. Translational Roadmap for rPPG Technology.

7. Conclusions

This review has synthesized the current landscape of rPPG and its viability in children with CP. It has been demonstrated that rPPG is a mature technology for HR monitoring. However, emotion detection remains a challenge due to the complexity of interpreting isolated physiological signals. In the pediatric population with CP, this challenge is further accentuated by spasticity and limitations in verbal communication, rendering conventional contact-based methods intrusive or ineffective.

It is necessary to improve performance in both clinical and domestic settings. A hybrid and multimodal methodological approach is recommended, that integrates algorithmic

robustness, sensor fusion (combining the rPPG signal with computer vision for body expressions and eye-tracking), and assistive technology.

These technologies empower clinicians by delivering an objective and proactive assessment of pain, stress, or well-being in non-verbal patients.

Currently, algorithms are primarily trained using healthy adult populations. Therefore, there is a need to expand the databases specifically for children with motor disabilities to train more inclusive and accurate algorithms, capable of taking into account the diversity of skin tones and movement patterns.

The transition toward telepsychiatry and remote monitoring systems will allow for the personalization of rehabilitation therapies in real-time, significantly reducing caregiver burden and improving communication between families and medical personnel.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app16115502/s1>; PRISMA 2020 Checklist [36].

Author Contributions: M.X.N.-B. investigation, formal analysis, writing—original draft preparation, writing—review and editing. V.H.C.-T. document review, content improvements, and editing. A.J.M.-C. document review, content improvements, and editing. I.M.G.-G. support in the literature search and review, development of the work structure, and review of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: This publication is part of the project PID2023–147508OB-I00, funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU awarded to BRETIA2 project.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

BCG	Ballistocardiography.
CHROM	Chrominance-based method.
CNN	Convolutional Neural Networks.
CP	Cerebral Palsy.
ECG	Electrocardiogram.
FPGA	Field Programmable Gate Array.
HR	Heart Rate.
HRV	Heart Rate Variability.
ICA	Independent Component Analysis.
LSTM	Long Short-Term Memory.
MAE	Mean Absolute Error.
PBV	Blood Volume Pulse Vector.
POS	Plane-Orthogonal-to-Skin.
RF	Respiration Frequency.
RMSE	Root Mean Square Error.
ROI	Regions Of Interest.
rPPG	remote PPG.
SNR	Signal-to-Noise Ratio.

References

1. Hsu, J.Y.; Jiang, T.Y.; Chao, P.C.P. A Fast FPGA Hardware Accelerator for Remote Heart Rate Detection Based on RGB Vision. *IEEE Trans. Biomed. Circuits Syst.* **2024**, *18*, 592–607. [\[CrossRef\]](#)
2. Liu, Y.; Qin, B.; Li, R.; Li, X.; Huang, A.; Liu, H.; Lv, Y.; Liu, M. Motion-Robust Multimodal Heart Rate Estimation Using BCG Fused Remote-PPG With Deep Facial ROI Tracker and Pose Constrained Kalman Filter. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5007215. [\[CrossRef\]](#)
3. Mehta, A.D.; Sharma, H. CPulse: Heart Rate Estimation From RGB Videos Under Realistic Conditions. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5023312. [\[CrossRef\]](#)
4. Sinhal, R.; Singh, K.; Raghuvanshi, M. Heart rate measurement based on color signal extraction. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 1990–1993. [\[CrossRef\]](#)
5. Qi, L.; Yu, H.; Xu, L.; Mpanda, R.; Greenwald, S. Robust heart-rate estimation from facial videos using Project ICA. *Physiol. Meas.* **2019**, *40*, 085007. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Wu, Y.C.; Lin, C.H.; Chiu, L.W.; Wu, B.F.; Chung, M.L.; Tang, S.C.; Sun, Y. Contact-Free Atrial Fibrillation Screening with Attention Network. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 5124–5135. [\[CrossRef\]](#)
7. Liu, H.; Ding, Y.; Zhou, M.; Li, Q. Adaptive-Weight Network for Imaging Photoplethysmography Signal Extraction and Heart Rate Estimation. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5023909. [\[CrossRef\]](#)
8. Gao, H.; Zhang, C.; Pei, S.; Wu, X. Region of Interest Analysis Using Delaunay Triangulation for Facial Video-Based Heart Rate Estimation. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5009712. [\[CrossRef\]](#)
9. Liu, S.Q.; Yuen, P.C. Robust Remote Photoplethysmography Estimation With Environmental Noise Disentanglement. *IEEE Trans. Image Process.* **2024**, *33*, 27–41. [\[CrossRef\]](#)
10. Collins, M.L.; Davies, T.C. Emotion differentiation through features of eye-tracking and pupil diameter for monitoring well-being. In Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 24–27 July 2023; pp. 1–4. [\[CrossRef\]](#)
11. Qi, Y.; Chee, Y.J.; Miao, C.; Zheng, S.; Jie Choo, T.W.; Zhang, R.; Wang, Q.; Qi Zhou, M.Y.; Olivo, M.; Dalan, R.; et al. An automated optical flow-mediated dilation method for fast screening of endothelial function. *Biomed. Signal Process. Control* **2026**, *118*, 109785. [\[CrossRef\]](#)
12. Liu, X.; Zhang, Y.; Yu, Z.; Lu, H.; Yue, H.; Yang, J. rPPG-MAE: Self-Supervised Pretraining With Masked Autoencoders for Remote Physiological Measurements. *IEEE Trans. Multimed.* **2024**, *26*, 7278–7293. [\[CrossRef\]](#)
13. Liu, L.; Xia, Z.; Zhang, X.; Feng, X.; Zhao, G. Illumination Variation-Resistant Network for Heart Rate Measurement by Exploring RGB and MSR Spaces. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5026613. [\[CrossRef\]](#)
14. Freitas, A.; Almeida, R.; Gonçalves, H.; Conceição, G.; Freitas, A. Monitoring fatigue and drowsiness in motor vehicle occupants using electrocardiogram and heart rate—A systematic review. *Transp. Res. Part F Traffic Psychol. Behav.* **2024**, *103*, 586–607. [\[CrossRef\]](#)
15. Boccignone, G.; Conte, D.; Cuculo, V.; D’Amelio, A.; Grossi, G.; Lanzarotti, R.; Mortara, E. pyVHR: A Python framework for remote photoplethysmography. *PeerJ Comput. Sci.* **2022**, *8*, e929. [\[CrossRef\]](#)
16. Gupta, K.; Sinhal, R.; Badhiye, S.S. Remote photoplethysmography-based human vital sign prediction using cyclical algorithm. *J. Biophotonics* **2024**, *17*, e202300286. [\[CrossRef\]](#)
17. Rosales, C.; Jácome, L.; Carrión, J.; Jaramillo, C.; Palma, M. Computer vision for detection of body expressions of children with cerebral palsy. In Proceedings of the 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), Salinas, Ecuador, 16–20 October 2017; pp. 1–6. [\[CrossRef\]](#)
18. Redd, C.B.; Silvera-Tawil, D.; Hopp, D.; Zandberg, D.; Martiniuk, A.; Dietrich, C.; Karunanithi, M.K. Physiological Signal Monitoring for Identification of Emotional Dysregulation in Children. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 4273–4277. [\[CrossRef\]](#)
19. Qiao, D.; Ayesha, A.H.; Zulkernine, F.; Jaffar, N.; Masroor, R. ReViSe: Remote Vital Signs Measurement Using Smartphone Camera. *IEEE Access* **2022**, *10*, 131656–131670. [\[CrossRef\]](#)
20. Liu, Y.; Xu, C.; Qi, L.; Li, Y. A robust non-contact heart rate estimation from facial video based on a non-parametric signal extraction model. *Biomed. Signal Process. Control* **2024**, *93*, 106186. [\[CrossRef\]](#)
21. Jo, J.; Yoon, Y.c. Remote Heart Rate Estimation Using Attention-targeted Self-Supervised Learning Methods. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2023**, *13*, 870. [\[CrossRef\]](#)
22. Benezeth, Y.; Krishnamoorthy, D.; Botina Monsalve, D.J.; Nakamura, K.; Gomez, R.; Mitéran, J. Video-based heart rate estimation from challenging scenarios using synthetic video generation. *Biomed. Signal Process. Control* **2024**, *96*, 106598. [\[CrossRef\]](#)
23. Wang, J.; Lu, H.; Wang, A.; Chen, Y.; He, D. Hierarchical Style-Aware Domain Generalization for Remote Physiological Measurement. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 1635–1643. [\[CrossRef\]](#)

24. Liu, X.; Yang, X.; Li, X. HRUNet: Assessing Uncertainty in Heart Rates Measured From Facial Videos. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 2955–2966. [[CrossRef](#)] [[PubMed](#)]
25. Das, M.; Choudhary, T.; Bhuyan, M.K.; Sharma, L.N. Non-Contact Heart Rate Measurement From Facial Video Data Using a 2D-VMD Scheme. *IEEE Sens. J.* **2022**, *22*, 11153–11161. [[CrossRef](#)]
26. Ouzar, Y.; Djeldjli, D.; Bousefsaf, F.; Maaoui, C. X-iPPGNet: A novel one stage deep learning architecture based on depthwise separable convolutions for video-based pulse rate estimation. *Comput. Biol. Med.* **2023**, *154*, 106592. [[CrossRef](#)]
27. Park, S.; Kim, B.K.; Dong, S.Y. Self-supervised RGB-NIR Fusion Video Vision Transformer Framework for rPPG Estimation. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5024910. [[CrossRef](#)]
28. Boccignone, G.; Conte, D.; Cuculo, V.; D’Amelio, A.; Grossi, G.; Lanzarotti, R. An Open Framework for Remote-PPG Methods and Their Assessment. *IEEE Access* **2020**, *8*, 216083–216103. [[CrossRef](#)]
29. Tangjui, N.; Taeprasartsit, P. Robust Method for Non-Contact Vital Sign Measurement in Videos Acquired in Real-World Light Settings From Skin Less Affected by Blood Perfusion. *IEEE Access* **2024**, *12*, 28582–28597. [[CrossRef](#)]
30. Shao, H.; Luo, L.; Qian, J.; Chen, S.; Hu, C.; Yang, J. TranPulse: Remote Photoplethysmography Estimation with Time-Varying Supervision to Disentangle Multiphysiologically Interference. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5029911. [[CrossRef](#)]
31. Kurihara, K.; Sugimura, D.; Hamamoto, T. Non-Contact Heart Rate Estimation via Adaptive RGB/NIR Signal Fusion. *Trans. Image Process.* **2021**, *30*, 6528–6543. [[CrossRef](#)]
32. Molinaro, N.; Zangarelli, F.; Schena, E.; Silvestri, S.; Massaroni, C. Cardiorespiratory Parameters Monitoring Through a Single Digital Camera in Real Scenarios: ROI Tracking and Motion Influence. *IEEE Sens. J.* **2023**, *23*, 20097–20106. [[CrossRef](#)]
33. Mirabet-Herranz, N.; Galdi, C.; Dugelay, J.L. Facial Biometrics in the Social Media Era: An In-Depth Analysis of the Challenge Posed by Beautification Filters. *IEEE Trans. Biom. Behav. Identity Sci.* **2024**, *7*, 108–117. [[CrossRef](#)]
34. Othman, W.; Kashevnik, A.; Ali, A.; Shilov, N.; Ryumin, D. Remote Heart Rate Estimation Based on Transformer with Multi-Skip Connection Decoder: Method and Evaluation in the Wild. *Sensors* **2024**, *24*, 775. [[CrossRef](#)]
35. Lysenko, S.; Seethapathi, N.; Prosser, L.; Kording, K.; Johnson, M.J. Towards Automated Emotion Classification of Atypically and Typically Developing Infants. In Proceedings of the 2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), New York, NY, USA, 29 November–1 December 2020; pp. 503–508. [[CrossRef](#)]
36. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.