Debnath and Kim
BioMedical Engineering OnLine (2025) 24:73
https://doi.org/10.1186/s12938-025-01405-5

BioMedical Engineering OnLine

REVIEW Open Access

A comprehensive review of heart rate measurement using remote photoplethysmography and deep learning

Uday Debnath[1] and Sungho Kim[1*]

*Correspondence:
sunghokim@yu.ac.kr

[1] Department of Electronic Engineering, Yeungnam University, 280 Daehak-Ro, Gyeongsan-Si 38541, Republic of Korea

Abstract

With the widespread availability of consumer-grade cameras, interest in heart rate (HR) measurement using remote photoplethysmography (rPPG) has grown significantly. rPPG is a noninvasive optical technique that uses camera to measure heart rate by analyzing light reflectance due to blood flow changes beneath the skin from any parts of the body, mostly facial regions. However, it faces challenges such as motion artifacts and sensitivity to varying lighting conditions. The rapid advancement of deep learning techniques in recent years has driven numerous studies to integrate these models with rPPG for HR detection in remote health monitoring systems. This study provides a comprehensive review of both conventional approaches and recent developments in rPPG and deep learning algorithms. A comparative analysis highlighted the superior accuracy of deep learning methods over conventional techniques in non-contact HR estimation. Based on a review of 145 articles encompassing different methodologies, signal processing strategies, and deep learning algorithms, our study identifies existing research gaps and explores future research opportunities for real-world applications.

Keywords: Heart rate measurement, RGB camera, Contactless monitoring, Remote photoplethysmography, Deep learning methods

Background

Heart rate (HR) is a critical physiological indicator essential for the early detection of cardiac arrhythmias and other cardiovascular conditions. Monitoring HR also aids in diagnosing various diseases and enabling timely treatment for improved patient outcomes [1]. Broadly, HR detection methods fall into contact-based and non-contact approaches.

Electrocardiography (ECG) is one of the most widely used contact-based techniques due to its accuracy and reliability [3]. It requires electrodes to be placed directly on the skin, which can potentially cause discomfort or allergic reactions [2]. In the Neonatal Intensive Care Unit, adhesive is used to connect sensors on the premature skin of babies which can lead to pain and skin irritation [2]. Additionally, ECG limits patient mobility, making it less convenient in certain scenarios such as continuous vital sign monitoring system. As an alternative, photoplethysmography (PPG) offers reliable HR measurements by mitigating these limitations [4].

Photoplethysmography (PPG) is an optical technique that monitors changes in blood volume within the microvascular bed of tissue by providing a more convenient and cost-effective alternative to ECG. PPG is generally less time consuming in terms of sensor setup and can obtain initial reading by using a single sensor whereas ECG requires electrodes properly placed on the body to measure heart rate. However, while the setup and initial acquisition with PPG can be faster, the overall duration of the measurement is ultimately driven by clinical requirements rather than by the technique itself. In other words, even though PPG may allow quick reading, if clinical conditions demand long-term monitoring, the monitoring period is set according to that need [81].

In the contact-based PPG method, a single sensor is positioned on the skin's surface usually on fingertip, earlobe, or other peripheral body region to detect changes in blood volume due to rhythmic fluctuations that align with physiological activities. The optical sensor contains a light source, a light-emitting diode (LED), and a photoelectrode. Unlike the single photodetectors employed in traditional PPG, remote photoplethysmography (rPPG) has the ability to take measurements over large areas of skin for increased signal-to-noise ratios and to measure pulse transit time by separating spatial information from the skin region. Light reflected from the skin consists of both specular and diffuse reflections; however, only the diffuse reflection contains pulsatile information relevant to rPPG [40]. The rPPG method uses different types of imaging devices such as webcams, smart-phone cameras, and industrial-grade cameras to assess multiple skin areas simultaneously. By illuminating the skin using the light source or LEDs and analyzing the changes in blood volume pulse on the skin area, physiological signs can be measured from the reflected photoplethysmographic transmission waveforms. These signal waveforms contain two key components: direct current (DC) and alternating current (AC). The DC component corresponds to the optical signal transmitted or reflected from the skin area, depending on the tissue structure as well as arterial and venous blood volume. The AC component fluctuates according to the change in blood volume during the cardiac cycle, while the DC component changes slowly with the respiration rate [126]. PPG signals can be affected by various factors such as wavelengths of light, motion artifacts, ambient light intensity, and temperature of the surrounding environment, whereas the rPPG signal is strongly affected by noise artifacts like the subject's motions, facial expressions, skin tone variation, and illumination conditions.

Beer-Lambert's law [5] states that light attenuation at a specific wavelength can be described in terms of the diffusion and scattering during light propagation through the skin. Previous studies [95] reveal that pulse oscillations of the arterial transmural pressure deform the connective tissue components of the dermis, resulting in periodic changes in light scattering and absorption. During the systole, the increased blood volume beneath the skin leads to increased light absorption. In contrast, during diastole, the decrease in blood volume reduces light absorption. Therefore, the light absorption strength by the skin is equally proportional to hemoglobin concentration level in blood and overall blood volumes in the tissues [95]. As hemoglobin absorbs green light better than red light and penetrates deeper than blue light, green light produces a pulsatile signal with a significant difference between the systolic and diastolic phases [93]. These variations, caused by blood circulation through the external carotid artery during heartbeats, result in subtle skin color changes that are imperceptible to the human eye but detectable with camera. rPPG signals can then be processed using blind source separation (BSS) techniques to extract HR [2]. After noise and motion artifacts are removed from the extracted rPPG signal, HR and heart rate variability (HRV) are determined by identifying peaks in the measured interbeat intervals.

Remote Photoplethysmography (rPPG) and Imaging Photoplethysmography (iPPG) are both non-invasive optical methods for detecting physiological signals such as heart rate and respiration rate by analyzing blood flow changes on the skin's surface; however, there are some differences between them. While iPPG and rPPG utilize the same basic input-video data, rPPG faces additional challenges in non-contact, uncontrolled scenarios. To consistently extract vital signs, a more complex method is required to address the existing challenges, such as motion artifacts and environmental conditions in the rPPG method. By using advanced deep learning approaches, the rPPG method can achieve robust ROI detection, reduce motion compensation, and improve overall signal quality. In the iPPG method, autoregressive (AR) spectral analysis is applied to extract the frequency bands of interest by analyzing the optical properties of skin color changes and motion-based methods. It is employed in controlled environments using imaging devices and sensors [93], with its main objective being to obtain the pulse waveform.

During the COVID-19 pandemic, remote health monitoring systems became essential worldwide. Social distancing and quarantine restricted movement, requiring individuals to wear masks in public. Consequently, many patients were unable to visit hospitals in person, creating a surge in demand for remote healthcare solutions through telemedicine and telehealth platforms [6, 7]. rPPG emerged as a valuable tool for monitoring vital signs on these platforms. Its non-invasive approach not only aids patient recovery but also reduces the risk of COVID-19 transmission among healthcare professionals [8]. As the pandemic ended in 2023, the post-pandemic era shifted the focus of rPPG research toward integrating this technology into remote health monitoring systems using low-cost cameras to monitor vital signs in resource-constrained environments, such as rural areas where access to healthcare facilities is very limited. Recent studies have shown that the rPPG method can be implemented for continuously monitoring patients remotely with chronic diseases such as hypertension and heart issues. Through virtual consultations, physicians can evaluate patient's health statuses via telehealth platforms [82]. rPPG can be utilized to assess heart rate and blood flow during sleep, offering significant insights into sleep quality and potential sleep problems such as insomnia and sleep apnea [83]. rPPG can also be implemented in pediatric settings, where non-invasive techniques are useful for monitoring the health conditions of infants and children [84].

Verkruysse et al. [9] first proposed the use of low-grade cameras with ambient light for con- tactless vital sign measurement. Since then, significant advancements have been made to address challenges such as motion artifacts and lighting variability, improving the accuracy of HR estima- tion. Although these traditional methods have made notable progress, they often rely on simplified assumptions that may not be applicable in real-world scenarios. Recent studies have demonstrated the potential of deep learning methods, which have achieved significant breakthroughs in applying digital signal processing for object detection [10] and image processing [11]. End-to-end and hy- brid deep learning models have consistently outperformed traditional computer vision techniques in HR estimation [28, 80].

Although there are many review articles on conventional and deep learning based heart rate detection using rPPG method [124, 125], these are insufficient to address latest progress in rPPG research and requires inclusion of latest studies to learn new methods and drawbacks in different scenarios to achieve further advancement in this field. The primary objective of this study is to provide a comprehensive review of recent advancements in noncontact HR detection, focusing on rPPG and deep learning methods applicable in different environments using RGB camera. By analyzing each neural network architecture, this study identifies existing challenges and proposes possible solutions to integrate rPPG methods for real-time applications. During the survey, this.

study have reviewed different datasets and listed.

The key contributions of this study include the following:

- *Systematic review of literature* A review of 145 relevant articles highlighting key findings , methodologies, and strategies for overcoming challenges in HR measurement using rPPG and deep learning approaches.
- *Methodology analysis* A comparative analysis of traditional and deep learning-based approaches to identify differences in architecture and performance for achieving superior accuracy in HR and HRV measurement systems.
- *Research applications and insights* A summary of recent advancements and potential applications in various fields, addressing challenges such as systematic biases and environmental conditions.
- *Limitations and future directions* Identification of key limitations and recommendations for introducing privacy protocols, improved algorithms, and enhanced reliability under diverse conditions for continuous HR monitoring in clinical and nonclinical settings.

This study is structured as follows: Sect."Materials and Methods" outlines the materials and methodology used for the article search and selection process, including the inclusion and exclusion criteria. Sect."Review Methodology" presents findings from the selected articles. Sect."Article Search Process" discusses various deep learning approaches and comparative analyses. Sect."Selection Process and Summary" explores limitations and future research opportunities. Finally, Sect."Results" concludes the study.

Materials and methods

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [12]. The following sections detail the search methodology of the study.

Review methodology

To systematically analyze recent advancements in rPPG methods, it was essential to first understand prior research in this domain. Information was gathered from leading digital libraries to obtain a comprehensive overview of the current state of the field. The research strategy focused on identifying relevant studies published between 2008 and early 2025.

Article search process

This study focuses on continuous, noncontact cardiac pulse estimation using rPPG with RGB cameras and deep learning algorithms. Relevant keywords were selected to guide the search process. These keywords include the following: "rPPG," "remote photoplethysmography," "remote," "deep learning," "heart rate," "heart rate measurement," "pulse rate information," "noncontact," "contactless," "RGB camera," "webcam," "camera," "facial images," "face video," "signal processing," "filtering," "noise reduction," "automatic cardiac pulse measurement," "RGB," "camera- based," "blood volume pulse," "pulse," "cardiac pulse estimation," "heart rate estimation," "physiological measurement," "physiological signals," "illumination variation," "motion artifacts," "dis- tance," "end-to-end," "convolutional neural network," "transformer network," "transformer-based architecture," "neural networks," "spatio-temporal network." Abbreviations for these terms were used where applicable to ensure comprehensive and efficient search.

Regarding the article search process, this study followed relevant keywords using AND and OR commands to search related articles from different databases. For example, to find relevant articles from IEEE Xplore we used the following search string: ((”remote photoplethysmography” OR”rPPG”) AND (”heart rate” OR”pulse rate”) AND (”RGB camera” OR”video” OR”camera” OR”facial video”) AND (”deep learning” OR”convolutional neural network” OR”CNN” OR”neural network”)). Table 1 provides an overview of article eligibility criteria.

Selection process and summary

Most of the articles for this review were sourced from Scopus, IEEE Xplore, Elsevier, Springer, and other reputable journals using Google Scholar and the defined keywords to identify a wide range of relevant publications. The selection process considered factors such as the title, complete abstracts and keywords, publication date, and journal type.

After compiling the initial list of articles, duplicates were removed. The selection process involved screening titles, abstracts, and full texts. During the title screening stage, studies unrelated to vital sign measurement, such as those focused solely on

Table 1 Article Eligibility Criteria

Inclusion Criteria	Exclusion Criteria
Published in English Peer-reviewed journal and conference articles Full-text availability Articles published between 2008 to early 2025	Studies in Medicine Business studies Agriculture and Biological Science Chemical Engineering
	Economics Pharmaceutical Science Microbiology and Biotechnology Fine Arts and Humanities Studies

face detection or facial expression anal- lysis, were excluded. Articles that used sensors other than RGB cameras, such as infrared, radar, LiDAR, and thermal sensors, were removed. This is because RGB cameras are the most available and accessible. Abstracts were then assessed against the same criteria, followed by full-text evaluations, to ensure adherence to the scope of HR measurement using rPPG and deep learning methods.

Key parameters considered during the selection process included population characteristics (e.g., age range, skin tone, and skin type), illumination conditions, camera specifications (e.g., frame rate, angular position, resolution), signal processing methods, BSS techniques, deep learning approaches, and performance metrics. An extensive search across databases, including Google Scholar, IEEE Xplore, and Springer, identified 2,290 related articles. A systematic screening and filtering process narrowed this number to 145 relevant articles for this review. Articles unrelated to HR measurement using rPPG and deep learning were excluded. Figure 1 provides a PRISMA workflow detailing the literature screening process.

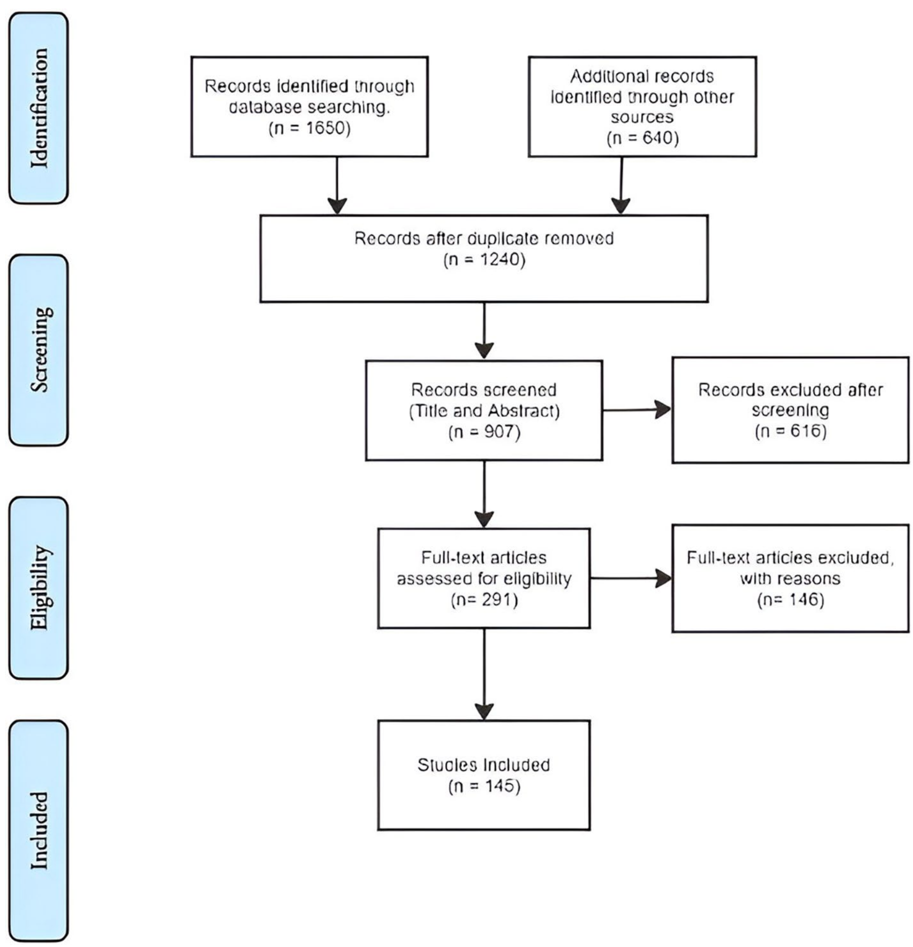

Fig. 1 PRISMA flow diagram for literature screening process

Results

In this section, we analyze the findings of our literature review. Throughout the article search process, we found that most articles were recently published in journals and conferences focusing on both rPPG and deep learning approaches. Figure 2 presents a statistical analysis of the number of articles published over the years. The data reveals a notable increase in interest after 2017, largely driven by the integration of deep learning algorithms with rPPG methods. This trend highlights the growing focus on advancing the methodology and addressing the challenges associated with rPPG for future research opportunities.

Existing datasets in rPPG research

There are several publicly available datasets that enable further research in vital sign measurement using noncontact sensors. These datasets facilitate the study of remote health monitoring methods, such as rPPG, and cover a range of environmental conditions and experimental setups. Table 2 provides an overview of publicly available datasets for rPPG research. Specifically, the MAHNOB-HCI dataset has become a benchmark for many researchers in emotion recognition and analysis through physiological signals [13]. It includes measurements of human emotions and physical reactions based on facial videos, ECG, EEG, respiration, and skin conductance. The study involved 27 participants (12 males and 15 females) aged 19 to 40 years, each participating in two sessions of approximately 30 min each. In this dataset, low-frequency signals are of particular interest for detecting HR or physiological signals using RGB sensors. These signals, typically within the range of 0.04–0.15 Hz, are especially relevant for HRV analysis and for identifying poor-quality signals. This dataset includes 256 Hz ECG dataset as reference data.

The PURE dataset includes 60 facial videos from 10 participants, with a distribution of 8 males and 2 females, aged 18 to 35 years [14]. Each participant contributed six

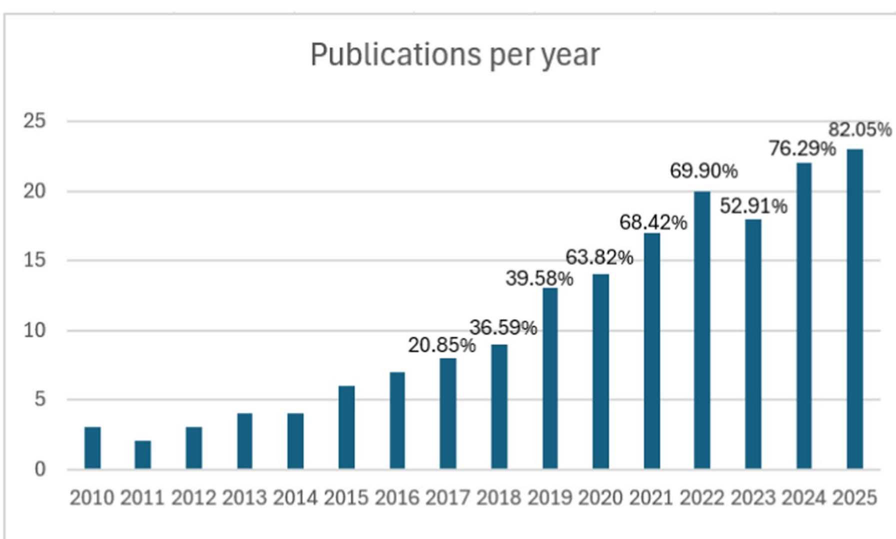

Fig. 2 Number of Publications per year and percentage of articles employing Deep Neural Network

Table 2 Available datasets for rPPG research

Database	No. of subjects	No. of videos	Age range	Camera details	Descriptions	GT Measurement
MAHNOB-HCI [13]	27	527	19–40 years old	Allied Vision Stingray F-046C; F-046B	61 fps, eye gaze, physiological sensors measuring ECG and EEG (32 channels), respiration amplitude and skin temperature	256 Hz ECG Ground Truth
PURE [14]	10	60	18–35 years old	evo274CVGE	30 fps, 640 × 480 resolution, Finger Pulse oximeter	CMS50E Pulse Oximeter
VIPL HR [15]	107	2378 (VIS),	22–41 years old	LogitechC310,	Visible light and near-infrared	CONTEC CMS60C BVP sensor
		752 (NIR)		Realsense F200, HUAWEI P9 smartphone	light videos, BVP-sensor data	
COHFACE [16]	40	160	20–50 years old	Logitech C525	20 fps, 640 × 480 pixels, heart rate and breathing rate	CMS50E Pulse oximeter
UBFC-rPPG [17, 18]	42	42	18–45 years old	Logitech C920 HD pro	30 fps, 640 × 480 pixels, CMS50E transmissive pulse oximeter	Pulse Oximeter
NBHR [19]	257	1130	New born babies 0–6 days old	Logitech C920 HD Pro	30 fps, PPG information, heart rate and oxygen saturation level	Pulse Oximeter
TokyoTech rPPG [20, 21]	9	9	20–29 years old	PrototypeRGB-NIR camera	300 fps, 640 × 480 pixels, contact PPG sensor	PPG Sensor
MMSE-HR [22]	40	102	18–40 years old	Di3D dynamic imaging system, FLIR A655sc	25 fps, 1040 × 1392 pixels, blood pressure signs	Biopac Mp150 BVP Sensor
MMPD [128]	33	660	25–35 years old	Samsung Galaxy S22 Ultra mobile phone	30 fps, 1280 × 720 pixels, Fitzpatrick skin types 3–6, four different lighting scenarios and 4 distinct activities	HKG-07C + Pulse Oximeter

videos, which captured typical movements such as quiet sitting, talking, and face rotation in a steady state under standard indoor lighting. While the skin tone categories are not explicitly coded, the dataset includes various skin tones. The goal of this dataset is to capture pulse rates within the normal frequency range of 0.75–2.5 Hz. However, the signals can be affected by motion artifacts, changes in illumination, or camera noise all of which may introduce low-frequency noise.

In more datasets, signals below 0.75 Hz, which typically fall outside the HR range, are filtered out during signal processing to improve the accuracy of pulse rate estimation. Using a CMS50E pulse oximeter attached to the subject's finger, the ground truth for PPG and SpO2 was acquired.

The VIPL-HR dataset includes 107 participants, although no specific distribution was provided. Participants' ages range from 22 to 41 years [15]. Recordings were made in various poses under different lighting conditions, including static poses, talking, and dynamic activities involving light motion. This dataset includes various facial movements and lighting environments, reflecting real-world scenarios. The ground truth heart rate in the VIPL-HR dataset was recorded using a CONTEC CMS60C BVP sensor.

The COHFACE dataset consists of facial videos from 40 participants aged 20 to 50 years [16], captured using similar recording settings as the VIPL-HR dataset. The UBFC-rPPG database contains videos from 42 participants aged 18 to 45 years, all in stable health. Each video is about 2 min long and was captured indoors under natural light. The NBHR database includes nearly 10 h of clinical video recordings of 257 newborn infants aged 0 to 6 days. Similarly as PURE dataset, CMS50E pulse oximeter is used to obtain reference rPPG dataset.

The TokyoTech rPPG dataset includes recordings from 10 subjects aged between 20 to 29 years old [20, 21]. The MMSE-HR dataset features 22 participants, with a balanced gender distribution and ages ranging from 18 to 60 years. These datasets were collected under controlled motion and lighting conditions. MMSE-HR dataset includes a total of 102 videos from 40 subjects whose age range is between 18 and 40 years old. The ground truth for the MMSE-HR dataset was measured using the Biopac Mp150 data acquisition system.

The MMPD [128] dataset includes 11 h of videos recorded using a Samsung Galaxy S22 ultra mobile of 33 subjects at a frame rate of 30fps. The ground truth PPG signals were simultaneously captured using an HKG-07C + oximeter. It contains subjects from different skin types, four different lighting scenarios such as LED-low, LED-high, incandescent, natural and subjects were performing four distinct activities like stationary, head rotation, talking, and walking as well as include exercise scenarios.

Most existing datasets use simpler methods for video collection. A variation that benefits categorization models, especially those intended for nonprofessionals, includes a diverse range of patient ethnicity, lighting conditions, experimental setups, and recording equipment. Most of the studies use public datasets to evaluate the deep learning model's performance.

In most studies, researchers used Pulse Oximeter for ground truth (GT) measurement due to its lower cost and wide availability [45]. Although ECG has been used in few studies and also provide golden standard accuracy while GT measurement [13], it has been less used due to its high price and complex settings. In 2008, Kalmin et al. [140] conducted a clinical study using a pulse oximeter to measure heart rate and compare its performance to ECG. This study shows that the pulse oximeter demonstrates reliable accuracy with a very low mean average error of 2 bpm compared to ECG when measuring heart rate in the delivery room. Therefore, it can be concluded that Pulse Oximeter is reliable as a verification method in rPPG method for HR measurement.

Conventional methods and signal processing approaches

In the rPPG method, BSS techniques are used to extract RGB signals by removing motion interference and systematic noise. These approaches assume that the pulse signal exhibits the strongest periodicity while other frequencies are excluded. ICA and PCA are the most commonly used BSS technique [13, 14]. Poh et al.[13, 14] applied the ICA method, noting that the green channel contains the strongest cardiac pulse–related information, as indicated in earlier studies by Verkruysse et al. They used the JADE algorithm, introduced by Cardoso et al. [44], to filter the raw RGB traces, eliminating motion artifacts due to BVP fluctuations. Lewandowska et al. [15] later proposed the PCA algorithm, which was implemented by Wang et al. [20]. The authors found that PCA generally provided better accuracy and higher computational speed than ICA. Recent studies have shown that rPPG signal extraction from a smaller rectangular region, such as the forehead, yields higher accuracy than using the entire face. However, both ICA and PCA methods showed lower accuracy when participants were physically active, and motion artifacts were present in the frames.

To overcome the motion sensitivity issue, the chrominance-based (CHROM) algorithm [17] was proposed. This algorithm models pulse information as a linear combination of RGB channels, assuming a standard skin tone. By utilizing color differences in the signals, CHROM reduces the effects of motion artifacts. Experimental results have demonstrated that CHROM is more robust to motion artifacts than other BSS techniques, particularly in motion periodicity. Table 3 presents an overview of various conventional signal-processing approaches.

Table 3 Signal processing approaches

Methods	Description
ICA [9, 13, 38, 41, 42, 47]	It decomposes the signal and derives independent pulse components from the temporal RGB data.
PCA [41–43, 48]	It is applied to reduce noise and derive key features and uncorrelated components from RGB traces.
CHROM [39, 41, 42, 45, 46, 49]	It is a linear combination of chrominance signals based on the assumption of skin color, and it isolates the pulse component while minimizing motion artifacts. However, it may not work when pulse signals coincide with specular reflections [64].
GREEN [40–42]	Hemoglobin and oxyhemoglobin in the blood absorb light at a peak wavelength of 520–580nm, which corresponds to the sensitivity of the green channel in standard RGB cameras [48, 65]. Due to its selective absorption, the green channel provides a greater signal-tonoise ratio for detecting physiological signals, particularly HR, than the red and blue channels.
POS [30, 40, 41]	This method involves projecting RGB derived signals onto an orthogonal plane of the temporally normalized skin tone vector, efficiently separating the small chromatic variations associated with physiological changes while limiting the impact of nonessential skin tone fluctuations.
SSR (Spatial Subspace Rotation) [42, 44]	This method extracts cardiac pulse information from the skin pixels and rotation measures. However, it may require a continuous sequence of camera frames to recover the pulse wave.

Basic workflow and experimental setup

This section reviews the basic workflow and experimental setup for contactless HR estimation using RGB cameras. Figure 3 provides a schematic overview of the experimental setup. For the experiment, which was conducted indoors, researchers used a ceiling lamp and a table lamp as two light sources to create different lighting conditions. The subject remained stationary, facing the camera, while an oximeter was used to obtain the ground truth for the PPG data. A webcam was used to record video data, which was synchronized with the oximeter for comparative analysis of the results. The subject was positioned 1 m away from the camera [60].

Figure 4 outlines the basic workflow for HR measurement using the rPPG technique and corresponding deep learning approaches for each stage. Previous studies have shown that, in most cases, subjects remain still while sitting in front of the camera to capture video [61]. The camera is used to capture facial images and color changes in the skin due to blood circulation. After capture, a face-tracking algorithm is applied to detect the ROI. Facial regions such as the fore- head and cheeks contain the most relevant pulse information for HR estimation [63]. Next, signal processing techniques are applied to extract the RGB channels from the frames. The green channel is often considered the strongest source of photoplethysmography signal, as it corresponds to the peak absorption of oxyhemoglobin. BSS techniques, such as independent component analysis (ICA) or principal component analysis (PCA), are used to isolate the signal from uncorrelated. It decomposes the signal and derives independent pulse components from the temporal RGB data. It is applied to reduce noise and derive key features and uncorrelated components from RGB traces.

It is a linear combination of chrominance sig- nals based on the assumption of skin color, and it isolates the pulse component while minimizing motion artifacts. However, it may not work when pulse signals coincide with specular reflections [62].

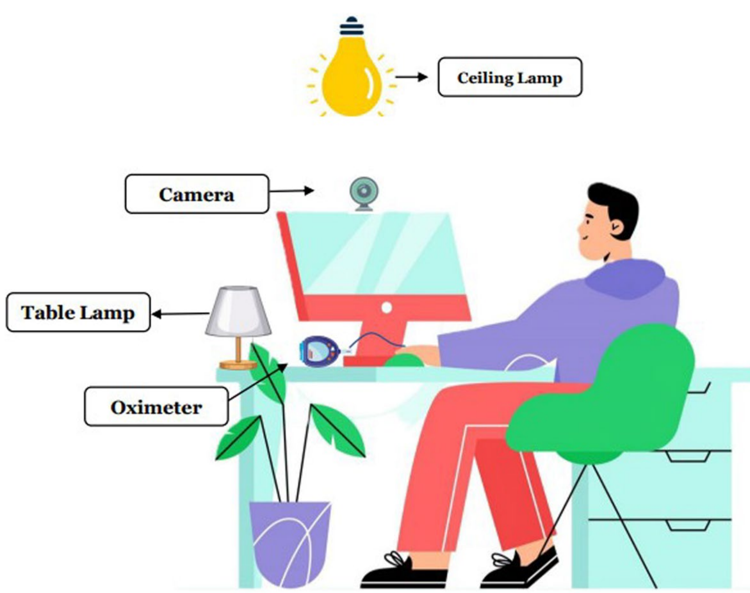

Fig. 3 A schematic representation of the experimental setup

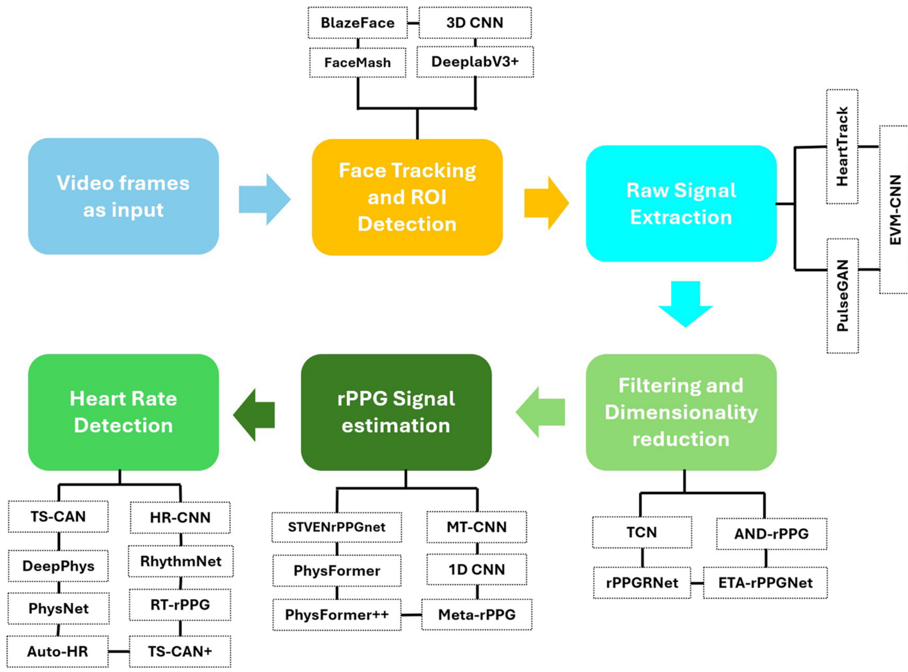

Fig. 4 Basic Workflow and corresponding Deep Learning Methods for HR Estimation

Hemoglobin and oxyhemoglobin in the blood absorb light at a peak wavelength of 520–580 nm, which corresponds to the sensitivity of the green channel in standard RGB cameras [47, 63]. Due to its selective absorption, the green channel provides a greater signal-to- noise ratio for detecting physiological signals, particularly HR, than the red and blue channels.

This method involves projecting RGB- derived signals onto an orthogonal plane of the temporally normalized skin tone vector, efficiently separating the small chromatic variations associated with physio- logical changes while limiting the impact of nonessential skin tone fluctuations.

This method extracts cardiac pulse information from the skin pixels and rotation measures. However, it may require a continuous sequence of camera frames to recover the pulse wave.

Components, as described in Sect. "Basic Workflow and Experimental setup". A band-pass filter is then applied to remove noise and unnecessary frequency components from the extracted signals. Typically, the frequency range for HR detection is between 0.75 Hz and 4 Hz, corresponding to an HR of 45 to 240 beats per minute (bpm) [38, 64]. Frequency domain analysis and peak detection algorithms are commonly used in conventional methods to measure HR from the extracted rPPG signals [65–67]. In the following sections, we have given an detailed overview of each of these stages. Table 4 highlights the overview of Non Deep Learning-based approaches for HR estimation using rPPG method.

Face tracking and ROI detection

Face tracking and Region of Interest (ROI) selection are among the most important stages for Blood Volume Pulse (BVP) signal extraction. In traditional methods,

Table 4 Non Deep Learning Methods for HR measurement using rPPG method

Ref	Proposed Algorithm	ROI Selection	Type of Feature	Basic Architecture	Results
[111]	SOBI	Eye Region	RGB Channel	ICA	Pearson Correlation Coefficient $r > 0.9$
[112]	ZCA	Nose, Forehead, Cheeks	RGB Channel	PCA	MAE = 1.15, Pearson Correlation Coefficient $r = 0.961$
[113]	Project ICA	Face Region	RGB Channel	ICA	RMSE = 7.21, Pearson Correlation Coefficient $r = 0.76$ (Stationary State)
[99]	KDICA	Both Cheeks	RGB Channel	CHROM, ICA	MAE = 1.26, Pearson Correlation Coefficient $r = 0.996$ (Distance 1 m)
[116]	GRGB	Whole Face Area	RGB Channel	G-R ratio, G-B ratio	Pearson Correlation Coeffi cient $r = 0.96$
[121]	TSS	Mouth, Nose	Green Channel	T-SNE	MAE = 1.08, RMSE = 4.04, Pearson Correlation Coefficient $r = 0.95$
[122]	Self-Adaptive SSA	Facial Skin Region	Green Channel	SSA	MAE = 2.05, RMSE = 5.27, Pearson Correlation Coefficient $r = 0.96$
[115]	FastICA	Forehead, Cheeks	Single Channel	ICA	MAE = 6.03, Pearson Correlation Coefficient $r = 0.96$
[117]	EEMD	Forehead, Cheeks	Green Channel	ICA, EMD	MAE = 3.812, Pearson Correlation Coefficient $r = 0.85$
[114]	U-LMA	Facial Skin Region	RGB Channel	ICA, LMA	RMSE = 3.85, Pearson Correlation Coefficient $r = 0.92$

the Viola–Jones algorithm is used to detect facial areas from the input frames [62]. In 2019, Project ICA [62] has been proposed to extract BVP signals using the skin reflection model. By incorporating feature point and pixel detection to track the skin area, it can overcome existing challenges caused by motion artifacts and noise in low-light conditions, improving the overall robustness of the model. Although it achieves significant performance improvement over conventional ICA, CHROM, SSR, and POS methods, Project ICA encounters challenges while extracting signals from dark skin variations [113]. On the other hand, deep learning models offer better precision by addressing the challenges of traditional methods. Deep learning models can be trained on diverse datasets to estimate HR by analyzing motion signals and rPPG signals in RGB video. For example, BlazeFace [143] is a face detection model developed by researchers at Google based on MobileNet network architecture.

Facemash is another widely used face detection model for precise ROI localization and feature extraction [118, 119]. These models eliminate redundant skin areas that do not impact the accurate detection of ROI.

Raw signal extraction

Independent Component Analysis (ICA) is widely used in conventional rPPG methods for blood volume pulse (BVP) signal extraction. It considers BVP extraction as a Blind Source Separation (BSS) technique and separates pulse related information by separating random noise and motion artifacts efficiently. Joint Approximate Diagonalization of Eigenmatrices (JADE) and FastICA algorithms enhance motion tolerance and improve the efficiency of ICA method while extracting pulse signals [13]. In 2017, Zhang et al., [111] proposed multi-channel ICA approach based on Second-order blind identification (SOBI) which can measure heart rate (HR) even in low illumination conditions.

On the other hand, deep learning-based approach leverages convolutional neural networks (CNNs) to extract accurate rPPG signals even from low-quality videos. Eulerian Video Magnification (EVM) allow accurate remote contactless HR estimation [27, 119]. Multi-task Temporal Shift Convolutional Attention Network (MTTS-CAN) facilitates contactless vital sign measurement by predicting both rPPG and respiratory signals [101]. However, this model requires complex preprocessing, which may limit its real-time application. In 2021, Song et al. [127], proposed PulseGAN, a generative adversarial network (GAN) that can denoise the signal waveform and significantly im- prove the overall model performance for HR estimation.

Signal filtering and rPPG signal estimation

In recent studies, authors showed that combination of different blind source separation (BSS) techniques can provide better accuracy to estimate HR in rPPG method. For example, Song et al. [99] proposed a Semi blind source separation (BSS) approach by combining ICA and CHROM method and separate independent features related to pulse information. However, this approach requires high-resolution videos to yield better performance. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) algorithm, combined with FastICA for remote HR measurement [123]. Due to residual white noise in this CEEMDAN algorithm, it causes decomposition errors that show lower performance in uncontrolled settings [123]. To address these challenges, Shi et al. [117] introduced an updated system by integrating zero-mean random white noise to sampled data and replacing the nonlinear function in FastICA with the Huber derivative approximation function to enhance overall model performance. Furthermore, Gupta et al. [114] proposed an under-complete ICA algorithm which preserves information across all RGB chan- nels by improving motion and illumination variation resistance while maintaining relatively higher accuracy.

Temporal and spatial features are the major factors while extracting rPPG signals from facial video [120]. In addition to processing the ST signals obtained by traditional methods, deep learning algorithms can integrate the Spatio-temporal (ST) module to improve model robustness by handling motion and noise artifacts. For example, ResNet-18 architecture is used to assess the quality of ST feature images extracted by the standard CHROM technique. This evaluation determines whether these feature characteristics should be used in the Fast Fourier Transform (FFT) for heart rate (HR) estimation.

Deep learning algorithms such as CNN-based end-to-end networks models have demonstrated higher accuracy in dynamic motion and illumination variations by processing spatio-temporal (ST) features from the input images to estimate HR. For example, DeeprPPG [59] is a lightweight CNN-based model that uses ST convolution network architecture to produce robust rPPG signals from input skin patches without need of a pre-processing requirement. Moreover, Hu et al. [49] proposed ETA-rPPGNet which includes a time-domain segment subnet and a backbone network. Time-domain segment subnet of this algorithm can significantly reduce redundant information of the feature maps extracted from video inputs. Also, integrated time domain attention mechanism in the backbone network can enhance the model's performance by handling noise artifacts efficiently.

Performance metrics

This section explains the most commonly used performance metrics and their formulas. We discuss various performance metrics to enable a comparative analysis of the proposed methodologies, focusing on the five most commonly used indicators from previous studies. In the reviewed articles, the effectiveness of each proposed method was evaluated using error performance indicators, such as root mean square error (RMSE), mean error (ME), standard deviation (SD), mean absolute error (MAE), and mean absolute percentage error (MAPE). Table 5 provides the formulas and descriptions of these metrics, which help assess the accuracy and reliability of methodologies used in non-contact heart monitoring systems. After reviewing numerous articles, we found that almost all used the MAE as the key performance indicator. Other studies also employed RMSE and the Pearson correlation coefficient (r) for comparative analysis. HR is typically measured in bpm to compare the estimated results from non-contact methods with the ground truth obtained from contact-based sensors.

Result analysis

In the reviewed articles, more than 90% used the MAE for comparison, while fewer than 65% used the RMSE. The unit of bpm was used to represent the difference between the estimated HR and the ground truth measured with contact-based sensors. Table 6 summarizes the performance metrics from the reviewed articles, detailing the methodologies and datasets used. The error metrics published across different studies show varying estimation errors, including ME, RMSE, MAE, and correlation coefficients (r). Most studies report low RMSE values, with the Deep-HR method achieving an RMSE

Table 5 Model performance metrics with formula

Metric	Formula	Description
Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{N}\sum_{t=1}^{N}(X_{Rlb}(t)-X_{Res}(t))^2}$	Root mean square error between the estimated and ground truth signals
Mean Error (ME)	$\frac{1}{N}\sum_{t=1}^{N}(X_{Rlb}(t)-X_{\mathrm{Res}}(t))$	Mean error between the estimated and ground truth signal
Mean Absolute Error (MAE)	$\frac{1}{N}\sum_{t=1}^{N}\lvert X_{Rlb}(t)-X_{\mathrm{Res}}(t)\rvert$	Mean absolute error between the estimated and ground truth signals
Mean Absolute Percentage (MAPE)	$\frac{1}{N}\sum_{t=1}^{N}\frac{\lvert X_{Rlb}(t)-X_{\mathrm{Res}}(t)\rvert}{\lvert X_{Rlb}(t)\rvert}\times 100$	Mean absolute percentage error between the estimated and ground truth signals
Standard Deviation (SD)	$\sqrt{\frac{1}{N}\sum_{t=1}^{N}(X_{Res}(t)-ME)^2}$	Standard deviation the estimated signal and ground truth

Table 6 Summary of papers on HR measurement using rPPG and Deep Learning Methods

Ref	Methodology	Type of Network	Dataset	Results	Experimental Setup
[23]	rPPG and Active Appearance models (AAM)	End-to-End	Custom Dataset	RMSE = 1.47	State = Stable, Indoor Condition, Natural Lighting condition
[24]	rPPG and HR-CNN	End-to-End	COHFACE, MANHOB-HCI, PURE, ECG Fitness Dataset	MAE = 1.84, RMSE = 2.37 (PURE)	State = Talking, Rowing, Exercising, Indoor Condition, 3 different light variation
[25]	Different rPPG technique	N/A	Custom Dataset	Limits of agreement: −8.3 to + 17.4	State = Sleeping Infants, Neonatal care Unit, Natural lighting condition
[26]	Image-based photoplethysmography	N/A	Custom Dataset	RMSE = 0.8887	State = Minimal movement, Indoor setup, Illumination = LED light on/off, camera flash
[27]	EVM-CNN	Hybrid	MMSE-HR	MAE = 6.85, RMSE = 6.95, Pearson Correlation coefficient (r) = 0.89	State = 2D Facial expression, Indoor setup, Ground Truth = HR, BP
[28]	DeepPhys	End-to-End	RGB I, RGB II and MAHNOB-HCI	MAE = 2.38, RMSE = 4.57	State = Sitting still, head motion, Indoor environment, Illumination = Natural and ambient light
[29]	rPPG and CNN	End-to-End	Welltory dataset	MAE (HR) = 5.91, MAE (SpO2) = 2.01, MAE (RR) = 3.11	State = eating, running, sleeping, walking, Indoor environment, Natural light
[30]	POS and 1D CNN	End-to-End	Custom Dataset	MAE = 2.78 bpm	State = Sitting still and stable, Indoor condition, General illumination condition
[31]	rPPG and CNN	End-to-End	MAHNOB-HCI, PURE	MAE = 3.12 bpm, SD = 3.78	State = still, talking, head movement, Indoor Environment, Frame rate = 60fps, Controlled and uncontrolled illumination from LCD screen
[32]	rPPG and Spatial–Temporal Attention Network	End-to-End	MMSE-HR, UBFC-rPPG	MAE = 0.23, RMSE = 0.48, Pearson Correlation coefficient (r) = 0.99	State = Controlled motion, sitting still and talking, frame rate = 20-25fps, Indoor condition, Natural and ambient light
[33]	rPPG with Structural Sparse Representation (SSR)	N/A	UBFC, COHFACE	MAE = 2.57, RMSE = 4.69	State = Various Poses, Indoor setting, Controlled and ambient lighting, Frame rate = 30fps, Ground truth = Heart rate, Breathing rate
[34]	rPPG and Singular Spectrum Analysis	N/A	UBFC-rPPG, CustomDataset	Pearson Correlation coefficient (r) = 0.89	State = Sitting still, Indoor settings,ambient illumination variation, Subject to webcam distance = 0.8 m

Table 6 (continued)

Ref	Methodology	Type of Network	Dataset	Results	Experimental Setup
[35]	rPPG and Spatial–Temporal Filtering	End-to-End	COHFACE, Custom Dataset	RMSE (Custom) = 0.82 RMSE (COHFACE) = 2.41	State = Sitting still without any body movements, Lab environment, Ambient light using 36 W tubelight in studio environment
[36]	Deep-HR	End-to-End	MAHNOB-HCI, HR-D (Custom Dataset)	MAE = 3.41 RMSE = 0.027	Subject sitting still watching movies, recorded different angle, Indoor environment, Normal lighting condition, Ground Truth = HR
[37]	SAMC	Conventional	MAHNOB-HCI, MMSE-HR	RMSE (MANHOB-HCI) = 6.23, RMSE (MMSE-HR) = 11.37	Subjects talking with various facial expression, Lab setting, Controlled and uncontrolled ambient lighting condition, Ground Truth = HR
[56]	STVEN-rPPGNet	Hybrid	OBF, MAHNOB-HCI	RMSE = 5.93 MAE = 4.03 Pearson Correlation Coefficient (*r*) = 0.88	Subject sitting still, Frame rate = 60fps, Lab setting, Ambient lighting using artificial light
[49]	ETA-rPPGNet	End-to-End	PURE, COHFACE, UBFC-rPPG	MAE = 1.46 RMSE = 3.97 (UBFC-rPPG), MAE = 2.66, RMSE = 6.48 (PURE)	State = Subject sitting still, talking, head rotation, Controlled Lab setting, Ambient and Natural lighting condition
[51]	AutoHR and NAS	End-to-End	MAHNOB-HCI, VIPL-HR	MAE = 3.78 RMSE = 5.10 Pearson Correlation Coefficient (r) = 0.72	State = Static poses, talking, facial movement and head motion, Indoor settings, different ambient lighting setup
[53]	HeartTrack (3D CNN)	End-to-End	MoLi-PPG-1, MoLi-PPG-2, UBFC-rPPG	MAE = 2.412, RMSE = 3.368 and Pearson Correlation Coefficient (r) = 0.983	Three different state = Static, Movement, Recovery after Physical Stress, Subject to camera distance = 1 m, ambient lighting condition using fluorescent light
[54]	Meta-rPPG (CNN-LSTM)	End-to-End	MAHNOB-HCI, UBFC-rPPG, Custom Dataset	MAE = 3.01, RMSE = 3.68, Pearson Correlation Coefficient (r) = 0.85	Subject playing time-sensitive mathematical game, Indoor setting, Ambient and Natural lighting condition
[52]	PhysNet	Hybrid	MAHNOB-HCI, OBF	MAE = 5.96 RMSE = 7.88 Pearson Correlation Coefficient (r) = 0.72	Static poses, talking, facial movement and head motion, Indoor settings, Controlled and uncontrolled lighting setup

Table 6 (continued)

Ref	Methodology	Type of Network	Dataset	Results	Experimental Setup
[131]	AND-rPPG	Hybrid	UBFC-rPPG, COHFACE	MAE = 3.15, RMSE = 4.75 (UBFC-rPPG), MAE = 7.83, RMSE = 8.06 (COHFACE)	Three different state = Static, Movement, recovery stage after exercise, Indoor Environment, Ambient lighting condition using fluorescent light on the ceiling
[133]	RTrPPG	End-to-End	VIPL-HR	MAE = 3.99	Three different state and Nine Scenario = Stable, motion, talking, ambient lighting (dark/bright), long distance, exercise, phone stable, and phone motion, Indoor settings, Natural and mbient lighting condition
[103]	Dual-GAN	Hybrid	PURE, UBFC-rPPG, VIPL-HR	MAE = 1.31, RMSE = 7.68 MAE = 0.82	State = Static, slight head rotation, Indoor setting, Ambient lighting condition
[78]	RADIANT	End-to-End	UBFC-rPPG, COHFACE, Synthetic dataset	MAE = 2.91, RMSE = 4.52	Subject sitting still, talking, head rotation, Controlled Lab environment, Ambient and Natural lighting condition
[79]	PhysFormer	End-to-End	MAHNOB-HCI, VIPL-HR, OBF, MMSE-HR	MAE = 2.84, RMSE = 5.36 (Cross-testing)	State = Static poses, talking, facial movement and head motion, Lab Environment, Controlled and ambient lighting condition
[80]	PhysFormer + +	End-to-End	MAHNOB-HCI, VIPL-HR, OBF, MMSE-HR	MAE = 2.71, RMSE = 5.15 (Cross-testing)	Static poses, talking, facial movement and head motion, Indoor settings, different ambient lighting setup
[101]	TS-CAN	End-to-End	AFRL, MMSE-HR	MAE = 3.41, RMSE = 7.82, Pearson Correlation Coefficient (r) = 0.82 (Cross-testing)	Subjects talking with various facial expression, Lab setting, Controlled and uncontrolled ambient lighting condition
[102]	TS-CAN +	End-to-End	PURE, UBFC-rPPG, MMPD	MAE = 1.31, RMSE = 2.66 (Intra-testing), MAE = 0.81, RMSE = 2.27 (Cross-testing)	Fitzpatrick skin types 3–6, four different lighting conditions (LED-low, LED-high, incandescent, natural), four distinct activities (stationary, head rotation, talking, and walking), and exercise scenarios

as low as 0.027. However, other statistical methods showed higher errors, particularly in MAE or correlation coefficient values, which help assess the efficiency of different algorithms applied to public or customized datasets. Among deep learning algorithms, There has been a growing trend of implementing more complex combinations of neural network architectures. Meta-rPPG (CNN-LSTM) generally shows the most balanced performance across several public datasets, including MAHNOB-HCI, VIPL-HR, and UBFC-rPPG. In contrast, Heart-Track achieves the lowest correlation coefficient error on the UBFC-rPPG dataset. Meta-rPPG demonstrated an MAE of 3.01 and an RMSE of 3.68, indicating reasonable accuracy and a good correlation coefficient (r =0.85). AutoHR, which uses advanced neural architecture search (NAS) for automated HR estimation, has higher errors (MAE =3.78, RMSE =5.10) but a slightly lower correlation coefficient error (r =0.77) compared to other models. In the VIPL-HR dataset, Physformer++ outperforms both AutoHR and Physformer at the lowest resolution settings. In Sect."Quantitative Performance Analysis of Model Performance on Public Dataset", we have performed a quantitative analysis of each model based on public datasets.

Research applications

Remote photoplethysmography (rPPG) has extensive applications in both clinical and nonclinical settings. Due to wide availability of smartphone devices, there have been ongoing studies for incorporating rPPG method into smartphone devices for non-contact physiological measurement. ReVise [104] is an end-to-end framework to measure physiological sign by using video of user's face captured using smartphone camera and rPPG method. Similarly, MobilePhys [106] trains on custom contactless camera-based PPG models by using a smartphone's front and rear cameras to provide high-quality self-supervised labels. At present, there is an ongoing trend of using wearable devices like smartwatches and smart bands. It is used by people from all ages as it allows users to continuously monitor their overall health conditions. Specially, Sportsperson and fitness trainers use these devices during practice and gym sessions. However, due to constant attachment of strips, it may cause skin irritation and allergic reaction if used for longer duration. So, Wang et al. [107], developed a non-contact rPPG method for HR detection during gym and fitness training sessions. Other studies have shown that, monitoring HR during sports activities can improve performance by providing instant feedback on athletes' vital signs [72, 73]. During the 2020 Tokyo Olympics, a contactless camera-based HR monitoring system was used to broadcast archers' HRs live on television. Four cameras, positioned 12 m away from the athletes, were equipped with pattern-recognition software to track facial color changes and calculate HR [75]. Recent studies concluded that the driver's fatigue and drowsiness due to long hours of driving is one of the main reasons behind car accidents [108]. In this case, rPPG can be used to improve overall driving safety by monitoring driver's overall health condition real-time. Huang et. al [71] suggested that integrating cameras into vehicles will enable us to monitor a driver's physiological status using rPPG method. Another study suggested using rPPG method to alert driver's abnormal vital signs and safely stop vehicles, reducing the risk of these accidents [71].

In clinical environments, rPPG is used to monitor patients' vital signs and assess their health continuously [69, 70]. Tran et al. [105], was first who demonstrated that it is possible to integrate rPPG method at a hospital. Also, camera-based HR monitoring systems can be implemented in Intensive Care Units during surgery or in post-anesthesia care to ensure critical patients are closely monitored [33]. It is also crucial to monitor newborns in Neonatal Intensive Care Units [25], where rPPG can provide noncontact vital sign measurements, reducing the risk of skin irritation from contact sensors while babies sleep [68, 69].

In 2019 [130], a clinical study was conducted by the researchers from the Neonatal Intensive Care Unit (NICU) at John Radcliffe Hospital in Oxford. Villarroel et al. recorded 426.6 h of video and reference vital signs from 30 preterm newborns over 90 sessions. Each preterm newborn was observed for up to 4 days in a row under normal ambient lighting condition during the day. Researchers created multi-task deep learning algorithms that could automatically separate skin areas and assess vital signs and no clinical interventions were required. The key objective of this study was to assess the accuracy heart rate and respiratory rate measurement on preterm newborn babies using video camera in clinical setting. Hatib et al., [129] performed a two phase clinical studies from January to April in 2023 on children using rPPG method for observing vital signs such as Heart rate, respiratory rate and oxygen saturation (SPO2) level in pediatric care unit and compared it to the current standard of methods. This study found that the use of rPPG technology is feasible and acceptable across varying age groups of paediatric patients, from neonates (up to 28 days old) to children aged between 5 and 16 years old. For older children aged 12 to 16 years old, they found good agreement between the rPPG technique when used to measure HR.

rPPG has been implemented during clinical surgeries as an intraoperative application. In this study [109], authors synchronized imaging photoplethysmography (iPPG) method with an electro- cardiogram (ECG) to act as a tool for monitoring tissue blood perfusion during abdominal surgery. Another study researcher developed an algorithm that implementing remote photoplethysmography (rPPG) method to use it as a continuous intraoperative and postoperative application for monitoring blood perfusion of the patients while undergoing reconstruction with free fasciocutaneous flaps (FFCF). Fasciocutaneous flaps, or axial flaps, consist of skin, subcutaneous tissue, and deep fascia, without muscular elements [144]. The main objective of the study was to determine if rPPG could provide reproducible information on flap perfusion and to define thresholds for its assessment [110].

Research outcome based on hardware specification

Most of the reviewed articles used RGB cameras [104, 106, 132, 137] due to its low cost and wide availability. There are several key parameters which may affect image quality captured by these cameras due to camera-subject distance, different spectrum range, sensor of the hardware and frame resolution. In this section, we will discuss these camera parameters and discuss about previous research outcomes based on hardware specifications in rPPG method.

Color format of video recording is an important factor for HR measurement using iPPG method. In this study [134], authors developed an HR measurement system that can handle both RGB and BayerBG color formats. They found that the BayerBG 8-bit color format reduces overall system error and provides optimal performance while estimating HR in iPPG method. Video compression parameters also have a significant interference in non-contact heart rate measurement. McDuff et al. [97] investigated the effects of video compression and found that it significantly reduced the quality of non-contact pulse signals. Zhao et al. [98] reviewed previous studies and proposed a novel rPPG signal-preserving framework to address this issue. Most of the studies considered camera frame rates between 10 and 30 fps while subjects are still stated. In terms of capturing fast-moving subjects, frame rates ranging from 100 to 500 fps were considered in this study [135]. However, Rapczynski et al. [100] concluded that videos captured using higher frame rates requires large storage space and increases overall computation cost.

There are different ranges of image resolution that has been observed among the reviewed articles but most of the studies captured subjects at a resolution 480 pixel to 720 pixel [136]. In deep learning based architectures, if the image resolution is too low, quantization noise interferes with the rPPG signal and becomes difficult to reduce the noise by spatial averaging from input video frames. Therefore, Lee et al., [96] proposed a 3D-CNN model which incorporates an attention module that would mainly focus on pulsatile signals in lower resolution images. It achieved similar accuracy across different image resolutions by incorporating adaptive lighting normalization and attention. This study shows that by integrating neural networks it can improve performance metrics through enhanced 3D CNN-based framework.

However, in another study [99], authors demonstrated that the image resolution significantly influences overall model performance, particularly when camera to subject distance exceeds 1 m. In further studies the multiple camera-setup, increased image resolution particularly at different camera to subject distances with diverse angles must be studied to setup camera for optimal model performance in rPPG method.

Color correction and Signal quality enhancement are crucial in rPPG methods to improve model's performance. Image enhancement methods (IEM) such as Gamma correction and Histogram equalization has been used to pre-process raw images and standardize them for rPPG signal extraction in conventional methods [139]. Chen et al. [138] proposed a deep learning-based image enhancement model (IEM), based on retinex theory to improve rPPG signal extraction and measured HR under different illuminations. This study showed that if PhysNet, a 3D convolution neural network (3D CNN) is applied without an IEM model, poor color correlation amplifies noise in the process and increases Mean Average Error by 60%.

Signal distortion is another parameter that affects the overall performance of the rPPG method. Therefore, Trumpp et al. [142] proposed a Camera-based Photoplethysmography (CbPPG) method by implementing polarization filtration technique to improve rPPG signal quality. They found that for an optimal illumination and perpendicular filter setting, their approach can improve signals' pulse strength significantly. Further studies should focus on developing optical filtering techniques and develop specified sensors using cross polarization approach to use on RGB camera to isolate specific wavelengths that contain pulse related information.

Result analysis of illumination variation

Illumination variation poses a significant challenge for remote PPG methods. Previous studies have explored different lighting conditions to optimize results. In one study [55], the authors ana- lyzed the performance of their algorithm for noncontact pulse measurement under varying lighting conditions and subject states. Each subject was recorded performing different tasks, including static conditions, head rotations (roll, yaw, and pitch), and body movements, such as walking to- ward the camera from a distance of 4 feet. During static conditions, with a light intensity of 150 lx, the RMSE was 5.56 bpm. At 300 lx, the performance showed slight variation, with an RMSE of 5.58 bpm. Under dynamic conditions, such as head rotations, mixed actions, and walking to- ward the camera, the RMSE values were 8.09 bpm, 9.96 bpm, and 8.60 bpm at 150 lx and 7.36 bpm, 9.88 bpm, and 7.72 bpm at 300 lx. These results indicate that RGB cameras perform worse under low light conditions. However, increasing illumination improves accuracy and reduces error in HR estimation.

Deep learning approaches

Deep learning-based methods are considered end-to-end processes because they automatically take video frames as input and estimate HR remotely. In the reviewed articles, several deep learning approaches have been implemented to improve the accuracy of HR detection. This section categorizes each method according to the types of networks used for automation. While some approaches involve preprocessing steps, such as ROI detection or estimating the rPPG (BVP) signal instead of directly estimating HR, these methods are still regarded as end-to-end frameworks be- cause they are trained using an end-to-end process. Figure 5 gives an overview of the deep learning approaches discussed in this article.

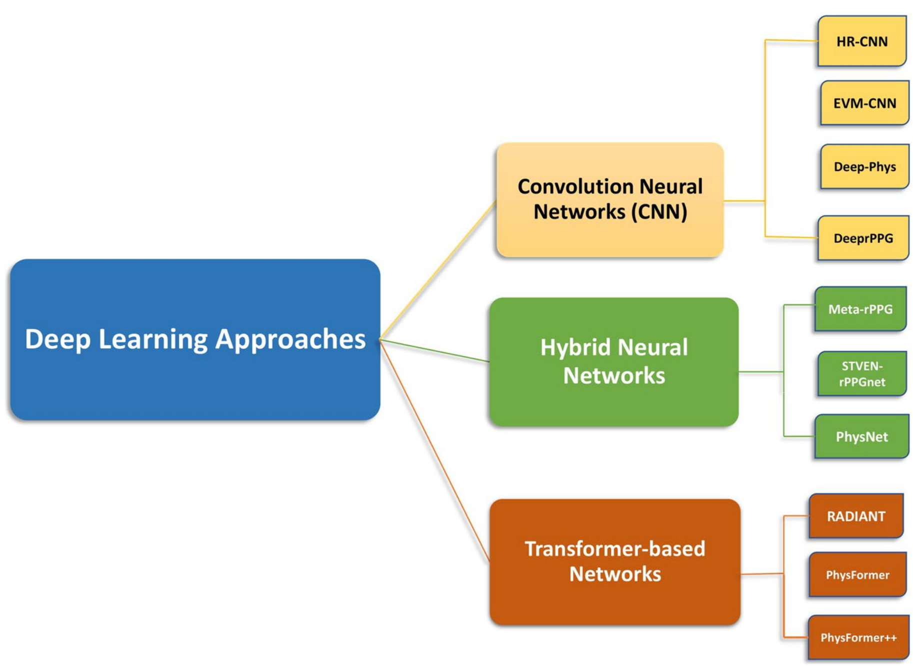

Fig. 5 An overview of deep learning approaches

Convolution neural networks (CNN)

Various convolutional neural network (CNN) architectures have been widely used in rPPG- based HR estimation. Some examples include convolutional attention networks (CANs), 2D CNNs, and 3D CNNs. These models process rPPG signals through hidden layers and perform feature extraction to estimate vital signs, such as HR. However, the limited availability of large datasets poses a challenge for effectively training deep learning models. To address this, researchers have used pretrained models trained on large image recognition datasets and other tasks. While conventional deep learning algorithms perform well on available datasets, some researchers have proposed custom CNN approaches for comparative analysis in combination with rPPG techniques.

One such approach is HR-CNN, a conceptual end-to-end method proposed by Spetlik et al. [49] for HR detection using rPPG. This model is based on a two-step 2D CNN architecture, which includes a signal extractor. The extractor is trained to generate rPPG signals from sequential inputs of the subject's facial images. The output signals are then processed through an estimator to measure HR. They trained their model on public datasets such as PURE, MAHNOB-HCI, and COHFACE and also conducted a comparative analysis by applying their algorithm to a custom dataset. The results showed that the proposed HR-CNN method outperforms traditional approaches, (e.g., CHROM and SSR) in terms of performance metrics such as RMSE, MAE, and Pearson correlation coefficient. While training and testing the model on PURE dataset, it achieved an MAE score of 1.84 and RMSE score of 2.37.

Another deep convolutional network, DeepPhys illustrated in Fig. 6, was proposed by Cheng et al. [50] as an end-to-end neural architecture for estimating both HR and breathing rates using the rPPG method. DeepPhys is based on two-branch 2D convolutional attention network archi- tecture (CAN) and can visualize the spatial–temporal distribution of physiological signs through its attention mechanism. The foundation of this model is a VGG-style CNN, which estimates physiological signals from motion representations. In this approach, the target signal is the first differential of the PPG waveform. It was trained and tested on Manhob-HCI dataset and showed a better performance with RMSE of 4.57 and MAE score of 2.38 bpm.

Liu et al. [59] introduced DeeprPPG, a hybrid deep learning technique combining spatio- temporal convolutional networks to process multiple video inputs of skin areas for HR estimation. During the COVID-19 pandemic, detecting HR from facial regions

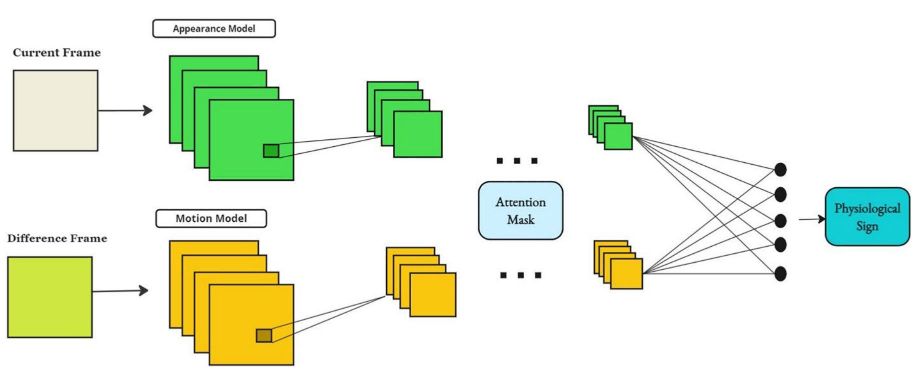

Fig. 6 DeePhys framework

became challenging when subjects wore masks. To address this, Zheng et al. [58] developed a CNN-based model capable of accurately estimating HR even when the subject is wearing a mask. This model demonstrated good performance in estimating HR under these conditions.

Combination of deep learning methods

Recent studies have seen the implementation of various deep learning approaches to improve results in rPPG detection. One notable algorithm, proposed by Lee et al., is Meta-rPPG [54]. This method introduces meta-based learning and transductive inference for rPPG detection. The model consists of a convolutional encoder, rPPG estimator, and synthetic gradient generator, combining CNN and LSTM methods. It performs well when trained on large datasets and adapts quickly when tested on new datasets. The synthetic gradient generator performs transductive inference and distance minimization on new video sequences, enabling rapid and supervised adaptation.

Another method, which was introduced in a previous study [56], combines rPPG with a spatio-temporal video-based estimation network (STVEN-rPPGnet). This approach merges two spatio- temporal networks (STNs) into a two-stage, end-to-end STN. A key challenge with this model is that it heavily compresses the input video frames, potentially removing spatially enhanced details. Despite this, the model can accurately detect HR and HRV using a partition constraint module and a skin-based attention module. An illustration of this deep learning approach is shown in Fig. 7.

Yu et al. [52] introduced the first end-to-end STN, PhysNet, to extract rPPG signal information from facial videos. This algorithm combines a 3D CNN, a 3D CNN with a temporal encoder and decoder, and a 2D CNN with an RNN for further compression. The model was trained and tested on the OBF and MAHNOB-HCI datasets. Their results showed that the combination of 3D CNN with a temporal encoder and decoder, named PhysNet64-3DCNN, provided the best accuracy for HR estimation compared to other methods. The authors demonstrated that this model outperforms HR-CNN and 2D-CNN-based approaches. Figure 8 shows an illustration of this deep learning approach from their study.

Neural architecture search

To enable continuous monitoring of cardiovascular and respiratory measurements, researchers are developing models to deploy rPPG for HR detection on mobile devices or in resource-constrained environments like homecare settings. Previous studies have shown that 3D CNNs outperform combinations of 2D CNNs and recurrent neural networks in terms of accuracy and performance [52]. NAS has been used to identify the most effective 3D CNN architecture for HR estimation, utilizing a loss function that

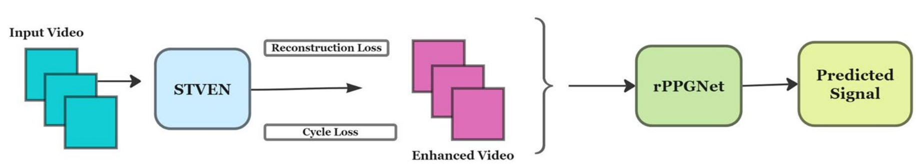

Fig. 7 Illustration of STVEN-rPPGnet framework

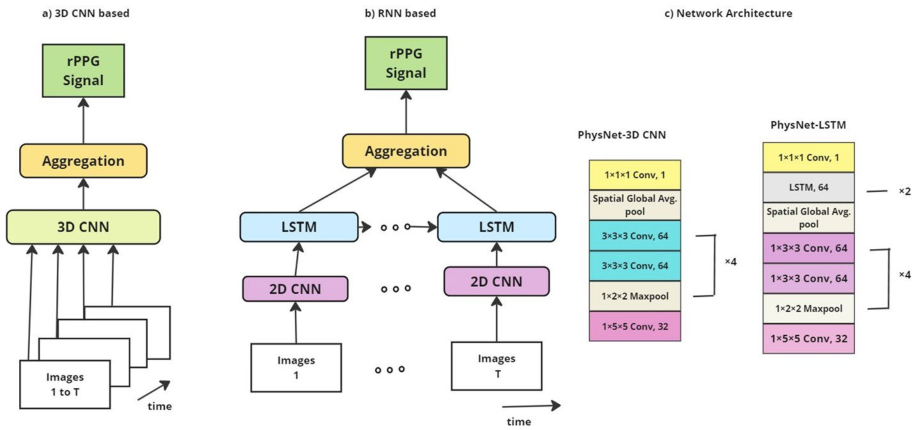

Fig. 8 The framework of PhysNet for rPPG signal recovery. **a)** 3D CNN based PhysNet; **b)** RNN based PhysNet; **c)** their corresponding network architectures

incorporates both time and frequency domain analysis. In 2020, Yu et al. [51] proposed an automated HR estimation system based on a novel approach called multitask temporal shift convolutional attention networks. This model was trained using the publicly available MAHNOB-HCI and VIPL-HR datasets.

Transformer-based network architecture

Transformer-based network architectures have been widely used in natural language processing tasks. Researchers from Google introduced the Vision Transformer (ViT) [76], a pure transformer model that can be applied directly to image sequences for classification. This innovation has significantly impacted the use of temporal architectures in computer vision applications. The primary goal of rPPG-based HR estimation approaches is to detect HR by filtering noise from multiple temporal channels and isolating pulse signals. Due to the strong correlation between temporal and estimated pulse signals, various transformer architectures can learn this relationship effectively. Yu et al. proposed an rPPG transformer framework called TransPPG [77], which is based on 3D mask face presentation attack detection (PAD) and facial recognition with background multiscale spatio-temporal maps. This model automatically extracts rPPG features by implementing the ViT architecture to separate pulse signals from temporal data for HR estimation. In 2022, Yu et al. [79] proposed PhysFormer, a transformer-based architecture that adapts both local and global temporal features to enhance rPPG representations. A key advantage of this model is its ability to train from scratch on rPPG datasets, unlike most transformer networks that require pretraining [79]. Recently, Yu et al. [80]. Upgraded PhysFormer to PhysFormer + +, incorporating slow-based temporal difference periodic and cross-attention transformers. In cross-sectional dataset testing, PhysFormer + + outperformed the original PhysFormer. In 2023, Gupta et al. [78] proposed RADIANT , a transformer-based pulse estimation method combining multilayer perceptron (MLP) and transformer architectures. MLP linear projection is used to obtain signal embeddings from temporal signals, improving rPPG feature representation, while the transformer architecture effectively denoizes the extracted signals and estimates cardiac pulse information. They demonstrated that their model performs better when pretrained on synthetic data.

Quantitative performance analysis of model performance on public dataset

In this section, we conducted a quantitative and comparative performance analysis of each deep neural network based on performance metrics evaluation for each dataset. From a comprehensive review of deep learning methodologies, it can be said that end-to-end CNN-based architectures were most commonly employed, and transformers and Generative Adversarial Networks (GANs) showed the best results across diverse dataset validations.

On the MAHNOB-HCI dataset, Physformer + + [80] outperforms PhysFormer [79] and achieved the lowest mean average error of 3.25 bpm and an RMSE score of 3.97 bpm, whereas HR-CNN showed the lowest correlation coefficient and a higher mean average error of 9.24 bpm during intra-dataset testing . On the COHFACE dataset, AND-rPPG [131] exhibits overall better performance compared to HR-CNN [24] and RADIANT [78], with a lower RMSE score of 8.06 bpm and a higher standard deviation of 0.92 bpm. On the PURE dataset, Dual-GAN [103] outperformed PulseGAN [127] and conventional methods with an RMSE score of 1.31 bpm and a mean average error (MAE) of.

0.82 bpm. Similarly, during testing on the UBFC-rPPG dataset, Dual-GAN [103] significantly outperforms other models with an RMSE score of 0.67 bpm and an MAE score of 0.44 bpm, whereas Meta-rPPG has comparatively higher error rates and a lower correlation coefficient score. On the VIPL-HR dataset, RTrPPG [133] demonstrates the lowest mean average error of 3.99 bpm and out- performs DualGAN and other deep learning algorithms.

While training on VIPL-HR and testing on the MMSE-HR (cross-testing), Physformer + + [80] surpasses its previous model Physformer and other deep learning models with the lowest mean average error of 2.71 bpm and an RMSE score of 5.15 bpm. TS-CAN + [102] outperforms TS-CAN with a 37.21% reduction in Mean average Error in cross-dataset testing on the UBFC-rPPG dataset and also achieved RMSE score of 2.27 bpm.

Discussion

In recent years, contactless HR measurement has gained significant attention, particularly with the advent of rPPG. The COVID-19 pandemic further accelerated advancements in rPPG techniques for using in remote health monitoring and telemedicine applications. However, several challenges remain in this field. In this section, we will discuss the current limitations and propose suggestions for future research.

Limitations

One major limitation is the lack of diversity in public datasets. Most datasets focus primarily on motion artifacts and illumination variations, which are key challenges in rPPG. High-quality and diverse open-source datasets are essential for benchmarking, analyzing, and developing more complex deep learning architectures. Existing datasets tend to lack demographic diversity, with most data being collected from individuals in the USA, Europe, or China, leading to limited representation of skin pigmentation.

Furthermore, the age range of participants in these datasets is typically narrow, with most subjects being young adults. To address these gaps, datasets should be expanded to include people from a broader range of ages and varied environmental conditions. Although extensive research has been conducted on illumination variation and motion artifacts, errors still occur in unconstrained environments. Other challenges, such as the distance between the camera and the subject, detection of multiple subjects under different lighting conditions, and varying camera specifications, have not been thoroughly addressed in many studies. As a result, models trained on these datasets may not predict HR accurately in real-world scenarios.

Data privacy is another critical concern. Since cameras are used to collect data for HR estimation, there is a risk of security breaches. Unauthorized entities, such as scammers, could misuse this data and steal sensitive information. Many studies focus only on facial regions for HR measurement, leaving other areas of the body underexplored. For instance, Gudi et al. [57] proposed a method for assessing psychological conditions through HRV, but this approach faced limitations when analyzing video frames with high compression. The use of spatial averaging techniques can also introduce noise, reducing the accuracy of HR detection.

Future research approaches

While the reviewed studies primarily focused on HR measurement, there is significant potential to detect additional physiological signals, such as blood pressure, oxygen saturation, and body temperature, using consumer-grade cameras and rPPG method. Heart rate variability or HRV is a crucial parameter for detecting the risk of heart failure and diagnosing breathing-related issues, which are linked to autonomic nervous system function.

Recently, there are few state of the art studies conducted on HRV measurement [85, 92], blood pressure [86, 87] and oxygen saturation (SPO2) [87] level measurement using rPPG method. However, authors concluded that proposed method [86, 87] has potential drawbacks due to data bias- ness and generalization as dataset only includes records from ICU patients and do not include any healthy subjects which limit these models applicability in real-world scenarios. By integrating transformers and attention-based modules and spatio-temporal modeling, deep neural architecture can improve overall model performance by mitigating noise artifacts which will overcome current drawbacks in conventional rPPG methods. Future studies should consider training and testing deep learning models on more diverse datasets incorporating detailed information such as skin tone variation, illumination variation, camera types, motion artifacts, demographic diversity, and age ranges with large sample sizes and long-duration of data.

Researchers should also explore the fusion of RGB cameras with other sensors, such as near-infrared, far-infrared, hyperspectral, or thermal sensors, to estimate a wider range of physiological signs and evaluate their performance in real-time scenarios. Some articles use data from multiple camera sources to overcome the restrictions imposed by a single data source. For example, Kurihara et al., [88] used RGB and NIR face patches to estimate the HR, which is useful in dark conditions and reduce background light. Jaiswal et al., [89] detected heart rate from facial video through RGB and MSR signal fusion.

Khanam et al. [90], detected breathing rate and and Negishi et al., [9] detected body temperature using RGB and thermal sensor fusion.

Given the growing concerns around data privacy, developing privacy protocols to prevent security breaches and address ethical concerns is essential. Future studies should also focus on applying the proposed methodologies in real-world clinical and nonclinical settings. rPPG and deep learning techniques can be used to enable biometric security facilities using face detection, in telehealth care facilities for virtual patient consultation, as well as in hospitals and home care settings for older patients. Future studies should be conducted for further improvement of image processing algorithm based on rPPG method for intraoperative applications during clinical surgery. Additionally, these techniques can be implemented in non-clinical environments (e.g., sports facilities, gyms, and fatigue monitoring systems) or in smartphone-based real-time vital sign monitoring applications.

Conclusion

This study provides a comprehensive review of conventional approaches and recent advancements in rPPG and deep learning methods for noncontact HR estimation using RGB cameras. We began by highlighting studies on experimental setups, signal processing methods, and feature extraction techniques applied to video images. In the results section, we analyzed existing databases and performance metrics based on parameters such as applied deep learning methods, subject movement, and dynamic illumination conditions across various public and custom datasets. Traditional methods often require face detection algorithms for ROI detection and various filters to extract cardiac pulse–related information. In contrast, we found that when deep learning algorithms are combined with rPPG, models can automatically detect the ROI and measure HR. These models require training on publicly available datasets. Hence, datasets need to be more diverse, including details such as age range, skin color, illumination variation, and movement to ensure performance across different demographics in real-time applications. In the discussion, we identified key challenges and limitations in this field and outlined future research directions and potential applications. Additionally, most articles we reviewed focused on short-duration HR monitoring within a limited distance. Long-term studies are needed to assess whether these technologies can maintain high accuracy over extended periods.

Abbreviations

BSS	Blind source separation
BVP	Blood volume pulse
LED	Light emitting diode
CAN	Convolutional attention networks
CNN	Convolutional neural network
HR	Heart rate
HRV	Heart rate variability
ICA	Independent component analysis
IRB	Institutional review board
MAE	Mean absolute error
MSR	Multi scale Retinex
ME	Mean error
MLP	Multilayer perceptron
PCA	Principal component analysis
NIR	Principal component analysis
PAD	Presentation attack detection

PRISMA	Preferred reporting items for systematic reviews and meta-analyses
RMSE	Root mean square error
ROI	Region of interest
SAMC	Self-adaptive matrix completion
SSR	Spatial subspace rotation
STN	Spatio-temporal networks
SOBI	Second-order Blind Identification
ZCA	Zero-phase component analysis
T-SNE	T-distributed stochastic neighbor embedding
EMD	Empirical mode decomposition
EEMD	Ensemble empirical mode decomposition
SSA	Singular spectrum analysis
TSS	T-SNE based signal separation
KDICA	Kernel density independent component analysis
CHROM	Chrominance
PCA	Principle component analysis
GRGB	The sum of green to red and green to blue channel
U-LMA	Unified Levenberg–Marquardt algorithm

Acknowledgements
This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (No. IRIS RS-2023-00219725). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (IRIS RS-2023-00240109).

Compliance with institutional review board regulations
In this review article, all information and data are collected from previously published studies. No new subjects were involved, and no additional data were collected specifically for this review. As such, this study meets the Institutional Review Board (IRB) regulations since it does not directly involve human or animal subjects. We have ensured that all cited research has followed ethical guidelines as reported by the original studies.

Author contributions
UD reviewed relevant studies, designed figures and tables, wrote statistical analysis, and drafted and revised the paper. SK reviewed the article and provided suggestion by analyzing the information and supervised throughout the study.

Funding
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. IRIS RS-2023-00219725). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (IRIS RS-2023-00240109).

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 27 November 2024 Accepted: 27 May 2025
Published online: 20 June 2025

References

1. Shelley KH. Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate. Anesth Analg. 2007;105:S31–6.
2. Challoner AV, Ramsay CA. A photoelectric plethysmograph for the measurement of cuta- neous blood flow. Phys Med Biol. 1974;19:317–28.
3. Singh A, Rehman SU, Yongchareon S, Chong PH. Multi-resident non-contact vital sign mon- itoring using radar: a review. IEEE Sens J. 2020;21:4061–84.
4. Johnston W, Mendelson Y. Extracting heart rate variability from a wearable reflectance pulse oximeter. In: Proceedings of the IEEE 31st Annual Northeast Bioengineering Conference, IEEE, 2005.
5. Swinehart DF. The Beer-Lambert law. J Chem Educ. 1962;39:333.
6. Watson AR, Wah R, Thamman R. The value of remote monitoring for the COVID-19 pan- demic. Telemed J E Health. 2020;26:1110–2.

7. Tang Z, Hu H, Xu C, Zhao K. Exploring an efficient remote biomedical signal monitoring framework for personal health in the COVID-19 pandemic. Int J Environ Res Public Health. 2021;18:9037.
8. Rohmetra H, Raghunath N, Narang P, Chamola V, Guizani M, Lakkaniga NR. AI-enabled remote monitoring of vital signs for COVID-19: methods, prospects and challenges. Com Puting. 2023;1:27.
9. Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. Opt Express. 2008;16:21434–45.
10. Fuertes D, Del-Blanco CR, Carballeira P, Jaureguizar F, Garcıa N. People detection with om- nidirectional cameras using a spatial grid of deep learning foveatic classifiers. Digit Signal Process. 2022;126:103473.
11. Wu W, Wang T, Wang Z, Cheng L, Wu H. Meta transfer learning-based super-resolution infrared imaging. Digit Signal Process. 2022;131: 103730.
12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, The PRISMA, et al. statement: an updated guideline for reporting systematic reviews. BMJ. 2020;2021:372.
13. Poh MZ, McDuff DJ, Picard RW. Advancements in noncontact, multiparameter physiologi- cal measurements using a webcam. IEEE Trans Biomed Eng. 2010;58(1):7–11.
14. Stricker R, Mu¨ller S, Gross HM. Non-contact video-based pulse rate measurement on a mo- bile service robot. In: The 23rd IEEE International Symposium on Robot and Human Inter- active Communication, IEEE, 2014.
15. Niu X, Han H, Shan S, Chen X. VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video. In: Comput Vis–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, Springer International Publishing, 2019.
16. Heusch G, Anjos A, Marcel S. A reproducible study on remote heart rate measurement. arXiv preprint arXiv:1709.00962. 2017
17. Bobbia S, Macwan R, Benezeth Y, Mansouri A, Dubois J. Unsupervised skin tissue segmen- tation for remote photoplethysmography. Pattern Recognit Lett. 2019;124:82–90.
18. Ryu J, Hong S, Liang S, Pak S, Chen Q, Yan S. Research on the combination of color chan- nels in heart rate measurement based on photoplethysmography imaging. J Biomed Opt. 2021;26: 025003.
19. Zhang L, Fu CH, Hong H, Xue B, Gu X, Zhu X, et al. Non-contact dual-modality emotion recognition system by CW radar and RGB camera. IEEE Sens J. 2021;21:23198–212.
20. Maki Y, Monno Y, Yoshizaki K, Tanaka M, Okutomi M. Inter-beat interval estimation from facial video based on reliability of BVP signals. In: 2019 41st Annual International Confer- ence of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019.
21. Xu G, Dong L, Yuan J, Zhao Y, Liu M, Hui M, et al. Rational selection of RGB channels for disease classification based on IPPG technology. Biomed Opt Express. 2022;13:1820–33.
22. Rapczynski M, Werner P, Al-Hamadi A. Effects of video encoding on camera-based heart rate estimation. IEEE Trans Biomed Eng. 2019;66:3360–70.
23. Datcu D, Cidota M, Lukosch S, Rothkrantz L. Noncontact automatic heart rate analysis in visible spectrum by specific face regions. In: Proceedings of the 14th International Confer- ence on Computer Systems and Technologies, 2013.
24. Sˇpetl´ık R, Franc V, Matas J. Visual heart rate estimation with convolutional neural network. In: Proceedings of the British Machine Vision Conference, Newcastle, UK 2018.
25. Gibson K, Al-Naji A, Fleet J, Steen M, Esterman A, Chahl J, et al. Non-contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: a method comparison study. Pediatr Res. 2019;86:738–41.
26. Sanyal S, Nundy KK. Algorithms for monitoring heart rate and respiratory rate from the video of a user's face. IEEE J Transl Eng Health Med. 2018;6:2700111.
27. Qiu Y, Liu Y, Arteaga-Falconi J, Dong H, El Saddik A. EVM-CNN: real-time contactless heart rate estimation from facial video. IEEE Trans Multimedia. 2018;21:1778–87.
28. Chen W, McDuff D. Deepphys: video-based physiological measurement using convolu- tional attention networks. In: Proceedings of the European con- Ference on computer vision (ECCV), 2018.
29. Mehmood A, Sarouji A, Rahman MM, Al-Naffouri TY. Your smartphone could act as a pulse- oximeter and as a single-lead ECG. Sci Rep. 2023;13:19277.
30. Firmansyah RA, Prabowo YA, Suheta T, Muharom S. Implementation of 1D convolutional neural network for improvement remote photoplethysmogra- phy measurement. Indones J Electr Eng Comput Sci. 2023;29:1326–35.
31. Liu L, Xia Z, Zhang X, Peng J, Feng X, Zhao G. Heart rate estimation by leveraging static and dynamic region weights. J Electron Imaging. 2023;32:043037.
32. Hu M, Guo D, Jiang M, Qian F, Wang X, Ren F. rPPG-based heart rate estimation using spatial-temporal attention network. IEEE Trans Cogn Dev Syst. 2021;14:1630–41.
33. Han J, Ou W, Xiong J, Feng S. Remote heart rate estimation by pulse signal reconstruction based on structural sparse representation. Electronics. 2022;11:3738.
34. Ryu J, Hong S, Liang S, Pak S, Chen Q, Yan S. A measurement of illumination variation- resistant noncontact heart rate based on the combination of singular spectrum analysis and sub-band method. Comput Methods Programs Biomed. 2021;200: 105824.
35. Das M, Choudhary T, Bhuyan MK, Sharma LN. Non-contact heart rate mea- surement from facial video data using a 2nd-vmd scheme. IEEE Sensors Jour Na L. 2022;22:11153–61.
36. Sabokrou M, Pourreza M, Li X, Fathy M, Zhao G. Deep-hr: fast heart rate estimation from face video under realistic conditions. Expert Syst Appl. 2021;186: 115596.
37. Tulyakov S, Alameda-Pineda X, Ricci E, Yin L, Cohn JF, Sebe N. Self-adaptive matrix com- pletion for heart rate estimation from face videos under realistic conditions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
38. Poh MZ, McDuff DJ, Picard RW. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express. 2010;18:10762–74.
39. De Haan G, Jeanne V. Robust pulse rate from chrominance-based rPPG. IEEE Trans Biomed Eng. 2013;60:2878–86.

40. Wang W, Den Brinker AC, Stuijk S, De Haan G. Algorithmic principles of remote ppg. IEEE Trans Biomed Eng. 2017;64:1479–91.
41. 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;8:216083–103.
42. Lewandowska M, Nowak J. Measuring pulse rate with a webcam. J Med Imaging Health Inform. 2012;2:87–92.
43. Huynh S, Balan RK, Ko J, Lee Y. VitaMon: measuring heart rate variability using smartphone front camera. In: Proceedings of the 17th Conference on Embedded Networked Sensor Sys- tems 2019.
44. Cardoso JF. High-order contrasts for independent component analysis. Neural Comput. 1999;11(1):157–92.
45. Bal U. Non-contact estimation of heart rate and oxygen saturation using ambient light. Biomed Opt Express. 2014;6(1):86–97.
46. Raseena KT, Ghosh PK. A maximum likelihood formulation to exploit heart rate variability for robust heart rate estimation from facial video. In: 2018 40th Annual International Confer- ence of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018.
47. Yu YP, Raveendran P, Lim CL, Kwan BH. Dynamic heart rate estimation using principal component analysis. Biomed Opt Express. 2015;6:4610–8.
48. Gutstein S. Remote photoplethysmography for cardiovascular monitoring. 2019.
49. Hu M, Qian F, Guo D, Wang X, He L, Ren F. ETA-rPPGNet: Effective time-domain atten- tion network for remote heart rate measurement. IEEE Trans Instrum Meas. 2021;12(70):1–2.
50. Chen W, McDuff D. Deepphys: video-based physiological measurement using convolu- tional attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018.
51. Yu Z, Li X, Niu X, Shi J, Zhao G. Autohr: a strong end-to-end baseline for remote heart rate measurement with neural searching. IEEE Signal Process Lett. 2020;27:1245–9.
52. Yu Z, Li X, Zhao G. Remote photoplethysmograph signal measure- ment from facial videos using spatio-temporal networks. arXiv preprint arXiv:1905.02419. 2019 May 7.
53. Perepelkina O, Artemyev M, Churikova M, Grinenko M. HeartTrack: convolutional neu- ral network for remote video-based heart rate monitoring. In-Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020.
54. Lee E, Chen E, Lee CY. Meta-rppg: remote heart rate estimation using a transductive meta- learner. Cham: Springer International Publishing; 2020.
55. Song R, Chen H, Cheng J, Li C, Liu Y, Chen X. PulseGAN: learning to generate realistic pulse waveforms in remote photoplethysmography. IEEE J Biomed Health Inform. 2021;25:1373–84.
56. Yu Z, Peng W, Li X, Hong X, Zhao G. Remote heart rate measurement from highly com- pressed facial videos: an end-to-end deep learning solution with video enhancement. In: Pro- ceedings of the IEEE/CVF International Conference on Computer Vision, 2019.
57. Gudi A, Bittner M, Lochmans R, Van Gemert J. Efficient real-time cam- era based estimation of heart rate and its variability. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
58. Zheng K, Ci K, Li H, Shao L, Sun G, Liu J, et al. Heart rate prediction from facial video with masks using eye location and corrected by convolutional neural networks. Biomed Signal Process Control. 2022;75: 103609.
59. Liu SQ, Yuen PC. A general remote photoplethysmography estimator with spatiotemporal convolutional network. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), IEEE. 2020.
60. Yang Z, Wang H, Lu F. Assessment of deep learning-based heart rate estimation using re- mote photoplethysmography under different illuminations. IEEE Trans Hum Mach Syst. 2022;52:1236–46.
61. Balakrishnan G, Durand F, Guttag J. Detecting pulse from head motions in video. In: Pro- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2013.
62. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2001.
63. Lee KZ, Hung PC, Tsai LW. Contact-free heart rate measurement using a camera. In: 2012 Ninth Conference on Computer and Robot Vision, IEEE. 2012.
64. Poh MZ, McDuff DJ, Picard RW. Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Transact Bio Med Eng. 2010;58:7–11.
65. Kwon S, Kim H, Park KS. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In: Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA. 2012.
66. Datcu D, Cidota M, Lukosch S, Rothkrantz L. Noncontact automatic heart rate analysis in visible spectrum by specific face regions. In: Proceedings of the 14th International Confer- ence on Computer Systems and Technologies 2013.
67. Holton BD, Mannapperuma K, Lesniewski PJ, Thomas JC. Signal recovery in imaging pho- toplethysmography. Physiol Meas. 2013;34:1499–511.
68. Villarroel M, Chaichulee S, Jorge J, Davis S, Green G, Arteta C, et al. Non-contact phys- iological monitoring of preterm infants in the neonatal intensive care unit. npj Digit Med. 2019;2:128.
69. Wieler ME, Murphy TG, Blecherman M, Mehta H, Bender GJ. Infant heart-rate measurement and oxygen desaturation detection with a digital video camera using imaging photoplethys- mography. J Perinatol. 2021;41:1725–31.
70. Huynh S, Balan RK, Ko J, Lee Y. VitaMon: measuring heart rate variability using smartphone front camera. In: Proceedings of the 17th Conference on Embedded Networked Sensor Sys- tems. 2019.
71. Wang J, Warnecke JM, Haghi M, Deserno TM. Unobtrusive health monitoring in private spaces: the smart vehicle. Sensors (Basel). 2020;20:2442.
72. Jimenez Morgan S, Molina Mora JA. Effect of heart rate variability biofeedback on sport performance, a systematic review. Appl Psychophysiol Biofeedback. 2017;42:235–45.
73. Bricout VA, DeChenaud S, Favre-Juvin A. Analyses of heart rate variability in young soccer players: the effects of sport activity. Auton Neurosci. 2010;154:112–6.

74. Huang PW, Wu BJ, Wu BF. A heart rate monitoring framework for real-world drivers using remote photoplethysmography. IEEE J Biomed Health Inf. 2020;25:1397–408.
75. Wells C. Archery debuts heart-rate graphics on broadcast of the olympic games.2021. https://worldarchery.sport/news/200395/ archery-debuts-heart-rate-graphics-broadcast-olympic-games
76. Dosovitskiy A. An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020.
77. Yu Z, Li X, Wang P, Zhao G. Transrppg: remote photoplethysmography transformer for 3d mask face presentation attack detection. IEEE Signal Process Lett. 2021;28:1290–4.
78. Gupta AK, Kumar R, Birla L, Gupta P. Radiant: better rppg estimation using signal embed- dings and transformer. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023
79. Yu Z, Shen Y, Shi J, Zhao H, Torr PH, Zhao G. Physformer: facial video- based physiolog- ical measurement with temporal difference transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022
80. Yu Z, Shen Y, Shi J, Zhao H, Cui Y, Zhang J, et al. PhysFormer++: facial video-based physiological measurement with slowfast temporal difference transformer. Int J Comput Vis. 2023;131:1307–30.
81. Rajala S, et al. Comparison of photoplethysmogram measured from wrist and finger and the effect of measurement location on pulse arrival time". Physiol Measure. 2018;39:075010.
82. McDuff D. Camera measurement of physiological vital signs. ACM Comput Surv. 2023;55:1–40.
83. Amelard R, Pfisterer KJ, Clausi DA, Wong A, Jagani S. Non-Contact As- sessment of Obstructive Sleep Apnea Cardiovascular Biomarkers Using Photoplethysmog- raphy Imaging. Opt Diag Sens XVIII: Toward Point-of-Care Diag. 2018;10501:1050113.
84. Gibson K, Al-Naji A, Fleet J, Steen M, Esterman A, Chahl J, et al. Non- contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: a method comparison study. Pediatr Res. 2019;86:738–41.
85. Gudi A, Bittner M, Lochmans R, Van Gemert J. Efficient real-time camera based estimation of heart rate and its variability. InProceedings of the IEEE/CVF international conference on computer vision workshops 2019 (pp. 0–0).
86. Omer OA, Salah M, Hassan L, Abdelreheem A, Hassan AM. Video-based beat-by-beat blood pressure monitoring via transfer deep-learning. Appl Intell. 2024;54(6):4564–84.
87. Chu Y, Tang K, Hsu YC, et al. Non-invasive arterial blood pressure measurement and SpO2 estimation using PPG signal: a deep learning framework. BMC Med Inform Decis Mak. 2023;23:131.
88. Kurihara K, Sugimura D, Hamamoto T. Non-contact heart rate estimation via adaptive rgb/nir signal fusion. IEEE Trans Image Process. 2021;14(30):6528–43.
89. Jaiswal KB, Meenpal T. rPPG-FuseNet: Non-contact heart rate estimation from facial video via RGB/MSR signal fusion. Biomed Signal Process Control. 2022;1(78): 104002.
90. Khanam FT, Perera AG, Al-Naji A, Mcintyre TD, Chahl J. Integrating RGB-thermal image sensors for non-contact automatic respiration rate monitoring. J Opt Soc Am A. 2024;41(6):1140–51.
91. Negishi T, Abe S, Matsui T, Liu H, Kurosawa M, Kirimoto T, Sun G. Contactless vital signs measurement system using RGB-thermal image sensors and its clinical screening test on patients with seasonal influenza. Sensors. 2020;20(8):2171.
92. Gudi A, Bittner M, Van Gemert J. Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation. Appl Sci. 2020;10(23):8630.
93. Matsumura K, Rolfe P, Lee J, Yamakoshi T. iPhone 4s photoplethysmography: which light color yields the most accurate heart rate and normalized pulse volume using the iPhysio Meter application in the presence of motion artifact? PLoS ONE. 2014;9(3): e91205.
94. Shao D, Yang Y, Liu C, Tsow F, Yu H, Tao N. Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time. IEEE Trans Biomed Eng. 2014;61(11):2760–7. https://doi.org/10.1109/tbme.2014.2327024.
95. Kamshilin AA, Nippolainen E, Sidorov IS, Vasilev PV, Erofeev NP, Podolian NP, Romashko RV. A new look at the essence of the imaging photoplethysmography. Sci Rep. 2015;5(1):10494.
96. Lee S, Lee M, Sim JY. DSE-NN: deeply supervised efficient neural network for real- time remote photoplethysmography. Bioengineering. 2023;10(12):1428.
97. McDuff DJ, Blackford EB, Estepp JR. The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. In2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) 2017. IEEE.
98. Zhao C, Chen W, Lin CL, Wu X. Physiological signal preserving video compression for remote photoplethysmography. IEEE Sens J. 2019;19(12):4537–48.
99. Song R, Zhang S, Cheng J, Li C, Chen X. New insights on super-high resolution for video- based heart rate estimation with a semi-blind source separation method. Comput Biol Med. 2020;1(116): 103535.
100. Rapczynski M, Werner P, Al-Hamadi A. Effects of video encoding on camera-based heart rate estimation. IEEE Trans Biomed Eng. 2019;66(12):3360–70.
101. Liu X, Fromm J, Patel S, McDuff D. Multi-task temporal shift attention networks for on- device contactless vitals measurement. Adv Neural Inf Process Syst. 2020;33:19400–11.
102. Li Y, Huang J, Zhao J, Wu D, Zheng M. TS-CAN+: A Improved TS-CAN Architecture for Non-Contact Heart Rate Measurement. IEEE Transactions on Consumer Electronics. 2024.
103. Lu H, Han H, Zhou SK. Dual-gan: Joint bvp and noise modeling for remote physiological measurement. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2021.
104. Qiao D, Ayesha AH, Zulkernine F, Jaffar N, Masroor R. Revise: remote vital signs mea- surement using smartphone camera. IEEE Access. 2022;15(10):131656–70.
105. Tran QV, Su SF, Sun W, Tran MQ. Adaptive pulsatile plane for robust noncontact heart rate monitoring. IEEE Trans Systems, Man Cyber: Syst. 2019;51(9):5587–99.
106. Liu X, Wang Y, Xie S, Zhang X, Ma Z, McDuff D, Patel S. Mobilephys: Personalized mo- bile camera-based contactless physiological sensing. Proc ACM Inter, Mobile, Wearable Ubiquitous Technol. 2022;6(1):1–23.

107. Wang W, den Brinker AC, Stuijk S, de Haan G. Robust heart rate from fitness videos. Phys- iological measurement. 2017;38(6):1023.
108. Goldenbeld C, van Schagen I, Davidse R. Fatigue-related consequences on road crashes. London: CRC Press; 2023.
109. Kamshilin AA, Zaytsev VV, Lodygin AV, Kashchenko VA. Imaging photoplethysmography as an easy-to-use tool for monitoring changes in tissue blood perfusion during abdominal surgery. Sci Rep. 2022;12(1):1143.
110. Schraven SP, Kossack B, Struder D, Jung M, Skopnik L, Gross J, Hilsmann A, Eisert P, Mlynski R, Wisotzky EL. Continuous intraoperative perfusion monitoring of free microvas- cular anastomosed fasciocutaneous flaps using remote photoplethysmography. Sci Rep. 2023;13(1):1532.
111. Zhang C, Wu X, Zhang L, He X, Lv Z. Simultaneous detection of blink and heart rate using multi-channel ICA from smart phone videos. Biomed Signal Process Control. 2017;1(33):189–200.
112. Iozzia L, Cerina L, Mainardi L. Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA. Physiol Meas. 2016;37(11):1934.
113. Qi L, Yu H, Xu L, Mpanda RS, Greenwald SE. Robust heart-rate estimation from facial videos using Project ICA. Physiol Meas. 2019;40(8): 085007.
114. Gupta A. A motion and illumination resistant non-contact method using undercomplete independent component analysis and Levenberg-Marquardt al- gorithm. IEEE J Biomed Health Inf. 2022;26(10):4837–48.
115. Favilla R, Zuccala VC, Coppini G. Heart rate and heart rate variability from single-channel video and ICA integration of multiple signals. IEEE J Biomed Health Infor- Matics. 2018;23(6):2398–408.
116. Haugg F, Elgendi M, Menon C. GRGB rPPG: An efficient low-complexity remote photoplethysmography-based algorithm for heart rate estimation. Bioengineering. 2023;10(2):243.
117. Shi Y, Qiu J, Peng L, Han P, Luo K, Liu D. A novel non-contact heart rate measurement method based on EEMD combined with FastICA. Physiol Meas. 2023;44(5): 055002.
118. Jewel JI, Hossain MM, Haque MD. Design and implementation of a drowsiness detection system up to extended head angle using facemesh machine learning solution. Cham: Springer Nature Switzerland; 2022.
119. Kolosov D, Kelefouras V, Kourtessis P, Mporas I. Contactless camera-based heart rate and respiratory rate monitoring using AI on hardware. Sensors. 2023;23(9):4550.
120. Yu Z, Li X, Zhao G. Facial-video-based physiological signal measurement: Recent advances and affective applications. IEEE Signal Process Mag. 2021;38(6):50–8.
121. Wang H, Yang X, Liu X, Wang D. Heart rate estimation from facial videos with motion interference using T-SNE-based signal separation. Biomed Opt Express. 2022;13(9):4494–509.
122. Wang D, Yang X, Liu X, Jing J, Fang S. Detail-preserving pulse wave extraction from facial videos using consumer-level camera. Biomed Opt Express. 2020;11(4):1876–91.
123. Lv W, Zhao Y, Zhang W, Liu W, Hu A, Miao J. Remote measurement of short-term heart rate with narrow beam millimeter wave radar. IEEE Access. 2021;10(9):165049–58.
124. Lee RJ, Sivakumar S, Lim KH. Review on remote heart rate measurements using photo- plethysmography. Mult Tools Applic. 2024;83(15):44699–728.
125. Pirzada P, Wilde A, Doherty GH, Harris-Birtill D. Remote photoplethysmography (rPPG): a state-of-the-art review. medRxiv. 2023;12:2023–110.
126. Tamura T. Current progress of photoplethysmography and SPO2 for health monitoring. Biomed Eng Lett. 2019;9(1):21–36.
127. Song R, Chen H, Cheng J, Li C, Liu Y, Chen X. PulseGAN: learning to generate realistic pulse waveforms in remote photoplethysmography. IEEE J Biomed Health Inform. 2021;25(5):1373–84.
128. Tang J, Chen K, Wang Y, Shi Y, Patel S, McDuff D, Liu X. Mmpd: Multi-domain mobile video physiology dataset. In2023 45th Annual International Conference of the IEEE Engi- neering in Medicine and Biology Society (EMBC) 2023. IEEE.
129. Hatib NA, Lee JH, Chong SL, Sng QW, Tan VS, Ong GY, Lim AM, Quek BH, How MS, Chan JM, Saffari SE. A two-phased study on the use of remote photoplethysmography (rPPG) in paediatric care. Ann Trans Med. 2024;12(3):46.
130. Villarroel M, Chaichulee S, Jorge J, Davis S, Green G, Arteta C, Zisserman A, McCormick K, Watkinson P, Tarassenko L. Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit. NPJ Digital Med. 2019;2(1):128.
131. Lokendra B, Puneet G. AND-rPPG: a novel denoising-rPPG network for improving remote heart rate estimation. Comput Biol Med. 2022;1(141): 105146.
132. Nadrag C, Poenaru V, Suciu G. Heart rate measurement using face detection in video. In2018 international conference on communications (COMM) 2018. 131–134. IEEE.
133. Botina-Monsalve D, Benezeth Y, Miteran J. RTrPPG: An ultra light 3DCNN for real-time remote photoplethysmography. InProceedings of the IEEE/CVF conference on computer vi- sion and pattern recognition 2022 (pp. 2146–2154).
134. Zhang C, Tian J, Li D, Hou X, Wang L. Comparative study on the effect of color spaces and color formats on heart rate measurement using the imaging photoplethysmography (IPPG) method. Technol Health Care. 2022;30(1):391–402.
135. Cerina L, Iozzia L, Mainardi L. Influence of acquisition frame-rate and video compres- sion techniques on pulse-rate variability estimation from vPPG signal. Biomedical Engineer- ing/Biomedizinische Technik. 2019;64(1):53–65.
136. McDuff DJ, Blackford EB, Estepp JR. Fusing partial camera signals for noncontact pulse rate variability measurement. IEEE Trans Biomed Eng. 2017;65(8):1725–39.
137. Gibson K, Al-Naji A, Fleet J, Steen M, Esterman A, Chahl J, Huynh J, Morris S. Non- contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: a method comparison study. Pediatr Res. 2019;86(6):738–41.
138. Chen S, Ho SK, Chin JW, Luo KH, Chan TT, So RH, Wong KL. Deep learning-based im- age enhancement for robust remote photoplethysmography in various illumination scenarios. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023. 6077–6085.

139. Yang Z, Wang H, Lu F. Assessment of deep learning-based heart rate estimation using re- mote photoplethysmography under different illuminations. IEEE Trans Human- Machine Syst. 2022;52(6):1236–46.
140. Kamlin CO, Dawson JA, O'Donnell CP, Morley CJ, Donath SM, Sekhon J, Davis PG. Ac- curacy of pulse oximetry measurement of heart rate of newborn infants in the delivery room. J Pediatr. 2008;152(6):756–60.
141. Kurita T, Kondo Y, Sun L, Moriuchi Y. Simultaneous acquisition of high quality rgb im- age and polarization information using a sparse polarization sensor. InProceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision 2023. 178–188.
142. Trumpp A, Bauer PL, Rasche S, Malberg H, Zaunseder S. The value of polarization in camera-based photoplethysmography. Biomed Opt Express. 2017;8(6):2822–34.
143. Bazarevsky V, Kartynnik Y, Vakunov A, Raveendran K, Grundmann M. Blazeface: Sub- millisecond neural face detection on mobile gpus. arXiv preprint arXiv:1907.05047. 2019 Jul 11.
144. Fix RJ, Vasconez LO. Fasciocutaneous flaps in reconstruction of the lower extremity. Clin Plast Surg. 1991;18(3):579–90.

Publisher's Note