electronics

MDPI

Article

Advancements in Remote Photoplethysmography

Linas Saikevičius [1,*], Vidas Raudonis [1], Agnė Kozlovskaja-Gumbrienė [2] and Gintarė Šakalytė [3]

1 Automation Department, Kaunas University of Technology, 44249 Kaunas, Lithuania; vidas.raudonis@ktu.lt
2 Department of Obstetrics and Gynecology, Lithuanian University of Health Sciences, 44307 Kaunas, Lithuania; agne.gumbriene@lsmuni.lt
3 Department of Cardiology, Medical Academy, Institute of Cardiology, Lithuanian University of Health Sciences, 44307 Kaunas, Lithuania; gintare.sakalyte@lsmu.lt
* Correspondence: linas.saikevicius@ktu.edu

Abstract: Advancements in camera technology over the past two decades have made image-based monitoring increasingly accessible for healthcare applications. Imaging photoplethysmography (iPPG) and remote photoplethysmography (rPPG) are non-invasive methods for measuring vital signs, such as heart rate, respiratory rate, oxygen saturation, and blood pressure, without physical contact. rPPG utilizes basic cameras to detect physiological changes, while rPPG enables remote monitoring by capturing subtle skin colour variations linked to blood flow. Various rPPG techniques, including colour-based, motion-based, multispectral, and depth-based approaches, enhance accuracy and resilience. These technologies are beneficial not only for healthcare but also for fitness tracking, stress management, and security systems, offering a promising future for contactless physiological monitoring. In this article, there is an overview of these methods and their uniqueness for use in remote photoplethysmography.

Keywords: rPPG; PPG; iPPG; photoplethysmography; remote sensors; vital signs; non-contact measurements

Academic Editor: Ping-Feng Pai

Received: 1 January 2025
Revised: 15 February 2025
Accepted: 26 February 2025
Published: 3 March 2025

Citation: Saikevičius, L.; Raudonis, V.; Kozlovskaja-Gumbrienė, A.; Šakalytė, G. Advancements in Remote Photoplethysmography. *Electronics* **2025**, *14*, 1015. https://doi.org/10.3390/electronics14051015

1. Introduction

In the past 20 years, advancements in camera technology have aligned with more accessibility and lower costs. This has sparked a growing interest in implementing these technologies in healthcare environments. Image-based monitoring technologies have the ability to measure numerous vital signs simultaneously utilizing a sensor that does not require physical contact. Imaging photoplethysmography is an optical method that uses a basic camera to evaluate various critical functions. Considerable efforts have been dedicated to accurately determining heart and respiratory rates. Presently, research is concentrating on remotely estimating pulse, respiratory rate, oxygen saturation, and blood pressure (BP). While there is an increasing amount of research on pulse and respiratory rate monitoring, there is a scarcity of publicly accessible papers on developments in blood pressure estimates.

Photoplethysmography (PPG) is changing the way we view health and wellness. Initially observed in 1936 by Alrick B. Hertzman [1], photoplethysmography found its first application in an oximeter invented by scientist Glen Millikan [2]. Since then, PPG has become increasingly common in devices, including oximeters of varying kinds and wearables, some of which are routinely used in hospitals today.

Photoplethysmography (PPG) is an affordable, non-invasive optical method utilized to measure changes in blood volume in peripheral tissues. A sensor is positioned on the skin's surface in PPG, usually on a fingertip, earlobe, or other peripheral body region. The sensor contains a light source, often a light-emitting diode (LED), and a photodetector. The light source delivers a beam of light into the tissue, and the photodetector calculates the light absorbed or reflected by the tissue.

Changes in blood volume as it circulates through tissues cause fluctuations in the tissue's light absorption or reflection. The changes are due to the pulsatile nature of blood flow, leading to regular variations in the intensity of the perceived light. These variations are referred to as photoplethysmographic waveforms.

The PPG waveform usually shows rhythmic fluctuations that align with physiological activities such as the heart cycle. During systole, as the heart pumps blood, the higher blood volume in the peripheral tissues causes increased light absorption or decreased reflection, leading to a downward deflection in the PPG waveform. During diastole, the relaxation phase of the heart, there is a reduction in blood volume, causing an upward deflection in the PPG waveform. PPG technology derives measurements by analyzing the variations in the waveform.

Remote photoplethysmography is a more sophisticated form of photoplethysmography that enables the collection of physiological data from a distance without the need for physical contact with the skin. PPG requires a sensor to touch the skin, such as on the fingertip, to monitor blood volume changes, whereas rPPG achieves this without direct contact. Remote photoplethysmography is a non-invasive imaging technique that allows for the evaluation of cardiovascular signals, like heart rate and blood flow, using just a camera. rPPG provides a convenient and contactless method of physiological monitoring by examining slight changes in skin colour due to fluctuations in blood volume. rPPG can be used for immediate monitoring of essential indicators, allowing for applications in healthcare environments, fitness monitoring, and stress control.

rPPG can be integrated into devices to provide intuitive interaction based on physiological responses, improving user experience in gaming, virtual reality, and emotion recognition systems. The distinct cardiovascular pattern detected by rPPG can be used for biometric authentication, improving security in access control systems.

2. Methods

Remote photoplethysmography refers to a variety of techniques used to monitor cardiovascular signals from a distance without invasive methods. The primary categories of rPPG include colour-based, motion-based, multispectral, depth-based, and mixed or hybrid approaches. A brief summary of each method suggested is provided below.

Colour-based approaches use variations in skin colour caused by blood circulation to determine physiological characteristics. GREEN (Plant-Orthogonal-to-Skin) employs ambient light to capture green channel data, while CHROM (chrominance-Based rPPG) concentrates on chrominance channels such as red and blue. Local Group Invariance (LGI) approaches utilize local colour changes, while the Blood Volume Pulse Signature (PBV) method analyzes blood volume pulse signals from colour channels. The POS (Algorithmic Principles of Remote PPG) approach utilizes colour information to predict heart rate, while OMIT (Face2PPG) extracts blood volume pulse from facial recordings in an unsupervised way.

Motion-based techniques utilize motion signals caused by blood flow to estimate heart rate. The independent component analysis method (ICA) isolates motion artefacts from rPPG signals, whereas Motion Magnification boosts slight movements in films to improve rPPG signals. Spatial and temporal filtering methods improve motion-related data, while deep learning methods like neural networks recognize motion patterns for rPPG estimates.

Multispectral techniques use many wavelengths of light, particularly near-infrared (NIR), to predict cardiovascular signals. NIR imaging records alterations in blood flow. Hemoglobin spectroscopy examines hemoglobin absorption spectra, while Dual-Wavelength Imaging utilizes dual-wavelength cameras to record rPPG signals.

Depth-based approaches utilize depth information acquired from depth sensors or 3D cameras. Depth PPG integrates depth and colour data, whereas 3D Convolutional Networks acquire depth-related characteristics to improve rPPG estimation accuracy.

Various methods are used together to incorporate colour, motion, and depth signals in order to improve the estimation of rPPG. Hybrid models combine data from several sources to enhance resilience and precision in physiological monitoring tasks.

Overall, rPPG methods consist of various strategies, each with its own principles, benefits, and uses. rPPG is evolving as a useful tool for non-invasive physiological monitoring across several fields by utilizing colour, motion, multispectral, and depth cues, together with improved signal processing and machine learning methods.

2.1. Colour-Based Methods in RPPG

Colour-based methods in rPPG leverage changes in skin colour captured by video data to extract cardiovascular signals, such as heart rate and blood volume pulse. By analyzing variations in skin chrominance or colour intensity over time, colour-based methods offer a non-invasive and contactless approach to physiological monitoring.

Key Components of Colour-Based Methods:

- Chrominance-Based Signal Extraction: Chrominance-based methods analyze colour information in video frames to extract cardiovascular signals. These methods often focus on chrominance channels, such as the Cb (blue-difference) and Cr (red-difference) channels in the YCbCr colour space, which are less affected by illumination changes and skin pigmentation compared to luminance channels.
- Spatial Averaging and Region of Interest (ROI) Tracking: Colour-based methods may involve spatial averaging techniques to aggregate colour information from multiple pixels within localized regions of interest (ROIs) on the skin. ROI tracking algorithms may also be employed to adaptively adjust the position and size of ROIs to track facial features or regions with optimal colour contrast.
- Frequency Domain Analysis: Colour-based methods often utilize frequency domain analysis to estimate cardiovascular signals from colour signals. Techniques such as Fourier analysis or spectral analysis can identify characteristic frequency components corresponding to heart rate or blood volume pulse in the frequency spectrum of colour signals.
- Machine Learning Integration: Machine learning algorithms may be integrated into colour-based rPPG methods for feature extraction, classification, or regression tasks. Supervised learning algorithms, such as support vector machines or neural networks, can learn patterns and relationships in colour data to improve the accuracy and robustness of cardiovascular signal extraction.

Robustness to Lighting Variations: Colour-based methods are less sensitive to changes in illumination compared to intensity-based methods, as they primarily focus on chrominance information. This robustness to lighting variations enables reliable signal extraction in diverse lighting conditions.

Suitability for Diverse Skin Tones: Colour-based methods can accommodate diverse skin tones and pigmentation levels, as they analyze chrominance information that is less influenced by melanin content compared to luminance information. This makes colour-based rPPG suitable for use across different populations and demographics.

Contactless Monitoring: Like other rPPG approaches, colour-based methods offer contactless and non-invasive monitoring of cardiovascular signals, eliminating the need for physical sensors or devices attached to the body.

Colour-based remote photoplethysmography techniques are utilized in various fields such as healthcare, wellness monitoring, biometrics, and human–computer interaction. They can be employed for the immediate evaluation of essential indicators, the identification of stress, the recognition of emotions, biometric verification, and interactive systems, among various other uses.

2.1.1. GREEN or Plant-Orthogonal-to-Skin

The GREEN (Greenness-Related Reflectance from rPPG and orthogonal-to-skin imaging) [3] method represents a pioneering advancement in remote photoplethysmography aimed at enhancing the robustness and accuracy of cardiovascular signal extraction. This paper presents a detailed overview of the GREEN method, highlighting its key components, advantages, and applications. By leveraging green light illumination and orthogonal-to-skin imaging, the GREEN method mitigates common challenges in rPPG, including motion artefacts, lighting variations, and skin tone differences. Through multispectral signal processing and temporal–spatial fusion techniques, the GREEN method achieves improved signal quality and reliability, enhancing the accuracy of physiological monitoring. With its versatility and applicability across diverse environments and populations, the GREEN method holds significant promise for various fields, including healthcare, sports science, and human–computer interaction.

The GREEN method integrates several innovative components to optimize cardiovascular signal extraction in rPPG. Green light illumination is employed for its unique properties, including deeper tissue penetration and reduced sensitivity to motion artefacts. Orthogonal-to-skin imaging, achieved by positioning the camera perpendicular to the skin surface, mitigates motion artefacts and enhances signal fidelity. Multispectral signal processing techniques enable the extraction of cardiovascular signals from green light reflectance captured by both conventional skin-facing cameras and orthogonal cameras. Additionally, temporal and spatial fusion techniques combine information from multiple imaging modalities, further enhancing signal quality and reliability.

The advantages of the GREEN method are as follows. Improved robustness: By leveraging green light and orthogonal-to-skin imaging, the GREEN method reduces sensitivity to motion artefacts, lighting variations, and skin tone differences, improving signal robustness in diverse environments and conditions. Enhanced accuracy: the GREEN method enhances the accuracy of cardiovascular signal extraction by minimizing noise and interference, resulting in more reliable measurements of heart rate, blood flow, and other physiological parameters, Versatility: green light illumination and orthogonal imaging make the GREEN method suitable for a wide range of applications and populations, including individuals with varying skin tones and levels of motion activity.

The GREEN method has applications in various fields, including healthcare, sports science, and human–computer interaction. It can be utilized for real-time monitoring of vital signs, stress assessment, biometric identification, and emotion recognition, among others.

2.1.2. CHROM

The CHROM [4] method represents an innovative technique in remote photoplethysmography designed to estimate physiological signals such as heart rate and blood flow by analyzing chrominance information extracted from colour video data. This paper provides a comprehensive overview of the CHROM method, discussing its key components, advantages, and applications. By focusing on chrominance channels and employing spatial averaging, temporal analysis, and signal processing algorithms, the CHROM method offers a contactless and robust approach to cardiovascular activity monitoring. Its wide applicability across various imaging modalities makes it suitable for diverse applications in healthcare, wellness monitoring, and human–computer interaction.

The CHROM method has emerged as a promising technique in remote rPPG for non-invasive monitoring of cardiovascular activity. By leveraging chrominance information extracted from colour video data, the CHROM method offers a contactless and robust approach to estimating physiological signals such as heart rate and blood flow. This paper provides an in-depth examination of the CHROM method, highlighting its key components, advantages, and applications in various domains.

The CHROM method relies on several key components to extract and analyze cardiovascular signals from chrominance channels. Chrominance extraction involves analyzing colour variations in the video data, primarily in the Cb (blue-difference) and Cr (red-difference) channels of the YCbCr colour space. Spatial averaging and filtering techniques are employed to enhance the signal-to-noise ratio and reduce the impact of motion artefacts. Temporal signal analysis tracks changes in chrominance values over time to estimate physiological signals, while signal processing algorithms such as Fourier analysis and independent component analysis (ICA) further enhance signal quality and mitigate noise artefacts.

The CHROM method offers several advantages over traditional rPPG approaches. Its contactless nature enables remote monitoring of cardiovascular activity without the need for physical sensors or devices attached to the body. By focusing on chrominance channels, which are less affected by motion artefacts compared to luminance channels, the CHROM method exhibits improved robustness in dynamic environments. Moreover, its wide applicability across various imaging modalities, including RGB cameras, thermal cameras, and hyperspectral imaging systems, enhances its versatility for different applications and scenarios.

The CHROM method is utilized in diverse fields such as healthcare, wellness monitoring, and human–computer interaction. It can be employed for immediate evaluation of essential signs, identification of stress, monitoring of sleep, and recognition of emotions, among various other uses. Due to its non-invasive nature and ability to withstand motion artefacts, it is highly suitable for dynamic environments and diverse populations.

2.1.3. LGI

The LGI [5] method offers an innovative approach to remote photoplethysmography by focusing on extracting cardiovascular signals through the exploitation of local spatial relationships in video data. This paper presents a detailed examination of the LGI method, elucidating its key components, advantages, and applications. By considering the temporal coherence of pixel groups within localized regions of interest, the LGI method aims to enhance signal robustness and reliability for physiological monitoring. Its robustness to motion artefacts, localized signal extraction capabilities, and adaptability to dynamic environments make it well suited for various applications in healthcare, fitness monitoring, and human–computer interaction.

The LGI method represents a novel approach to rPPG aimed at improving the robustness and reliability of cardiovascular signal extraction. By exploiting local spatial relationships in video data, the LGI method offers a promising solution to the challenges posed by motion artefacts and dynamic environments. This paper provides an in-depth analysis of the LGI method, highlighting its key components, advantages, and applications across various domains.

The LGI method leverages several key components to extract cardiovascular signals from video data. It focuses on local spatial relationships among pixels within regions of interest, considering the collective behaviour of pixel groups with similar motion and colour characteristics. By analyzing pixel groups, the LGI method exploits the invariance properties of local motion and colour patterns over time, mitigating the effects of motion artefacts and noise. Temporal signal analysis techniques track changes in colour and motion signals within each group over successive frames, enabling the estimation of cardiovascular parameters. Adaptive signal processing techniques, including Spatial Filtering and frequency domain analysis, further enhance signal quality and robustness.

The LGI method offers several distinct advantages over traditional rPPG approaches. Its focus on local spatial relationships improves robustness to motion artefacts and camera motion, as it analyzes coherent pixel groups within ROIs. Furthermore, the LGI method enables localized extraction of cardiovascular signals, allowing for targeted analysis of specific regions of interest within the video data. Its adaptability to dynamic environments and varying lighting conditions further enhances its utility in real-world scenarios.

The LGI method is utilized in diverse fields such as healthcare, fitness monitoring, and human–computer interaction. It can be used to promptly evaluate vital signs, detect stress, track activity, and recognize emotions in various environments. The tool's strength, ability to extract signals in specific areas, and flexibility make it highly valuable for monitoring physiological changes in dynamic environments.

2.1.4. PBV Method

The PBV [6] method represents an innovative technique in rPPG aimed at extracting cardiovascular signals based on changes in blood volume in the skin. Unlike traditional rPPG methods that focus on colour or motion variations, the PBV method directly estimates blood volume pulse signals from video data, offering a robust and direct measure of cardiovascular activity. This paper provides a detailed overview of the PBV method, discussing its key components, advantages, and applications. By analyzing intensity variations in video frames, employing spatial and temporal filtering techniques, and extracting relevant features, the PBV method enables real-time monitoring of blood volume pulse signals, making it suitable for diverse applications in healthcare, wellness monitoring, and sports science.

The Blood Volume Pulse Signature method offers a novel approach to photoplethysmography by directly estimating blood volume pulse signals from video data. This paper provides a comprehensive examination of the PBV method, highlighting its key components, advantages, and applications. By focusing on intensity variations related to blood volume changes and employing robust signal processing techniques, the PBV method enables real-time monitoring of cardiovascular activity, making it a valuable tool for diverse applications in healthcare, wellness monitoring, and sports science.

The PBV method relies on several key components to extract cardiovascular signals from video data. It estimates changes in blood volume in the skin by analyzing intensity variations in video frames, providing a direct measure of cardiovascular activity. Spatial and temporal filtering techniques are employed to enhance the accuracy and reliability of blood volume estimation while suppressing noise and artefacts. Feature extraction methods extract relevant characteristics of the blood volume pulse signals, such as peak amplitudes and waveform morphology, for further analysis and interpretation. Calibration and validation steps may be required to establish a quantitative relationship between extracted signals and physiological parameters, ensuring the accuracy and consistency of PBV measurements across different individuals and conditions.

The PBV method offers several distinct advantages over traditional rPPG approaches. Its direct estimation of blood volume pulse signals provides a reliable measure of cardiovascular activity without relying on secondary indicators such as colour or motion variations. By focusing on intensity variations related to blood volume changes, the PBV method exhibits improved robustness to motion artefacts and camera motion, ensuring accurate measurements in dynamic environments. Furthermore, its ability to enable real-time monitoring from standard video sources allows for continuous and non-invasive assessment of cardiovascular activity in various applications and environments.

The PBV method finds applications across diverse domains, including healthcare, wellness monitoring, and sports science. It can be utilized for real-time assessment of vital signs, stress detection, fatigue monitoring, and performance evaluation in diverse populations and settings. Its direct estimation of blood volume pulse signals and robustness to motion artefacts make it particularly well suited for dynamic environments and real-time monitoring scenarios.

2.1.5. POS

The POS [7] method presents a comprehensive framework for remote photoplethysmography, encompassing a set of algorithmic principles and techniques for signal acquisition, processing, and analysis. Its overarching objective is to extract accurate and reliable cardiovascular signals from video data captured by standard cameras. The POS method integrates various computational approaches to achieve this goal.

One fundamental aspect of the POS method is motion compensation, which addresses the challenge of motion artefacts caused by subject movement or camera motion. By employing motion estimation and compensation algorithms, the POS method aligns video frames to minimize distortions in the photoplethysmographic signals. Additionally, spatial and temporal filtering techniques are utilized to enhance signal quality and reduce noise. Spatial filters such as Gaussian filters and median filters smooth pixel intensity variations, while temporal filters like moving averages and Kalman filters remove high-frequency noise and artefacts from the temporal signal.

Another key component of the POS method is feature extraction and selection, which involves identifying relevant physiological features from the video data. These features encompass colour-based metrics, motion characteristics, frequency domain parameters, and statistical measures derived from the photoplethysmographic signals. Feature selection methods, such as principal component analysis and recursive feature elimination, assist in identifying the most informative features for signal analysis. Furthermore, signal decomposition and reconstruction techniques are employed to separate and extract individual physiological components when multiple signals are present in the video data. Methods such as independent component analysis and wavelet decomposition aid in isolating specific physiological signals, such as heart rate or respiratory rate.

Machine learning integration is also integral to the POS method, facilitating tasks such as signal classification, regression, and anomaly detection. Supervised learning algorithms, including support vector machines, random forests, and neural networks, are trained on labelled datasets to classify physiological states or predict clinically relevant outcomes based on extracted features.

The POS method provides numerous benefits, such as flexibility, durability, and precision. The versatility of this technology in accommodating different imaging modalities, camera configurations, and environmental circumstances renders it highly applicable for a diverse array of remote photoplethysmography applications. The POS method exhibits enhanced resilience to motion artefacts, lighting variations, and other forms of noise in the video data by integrating motion compensation and noise reduction techniques. The POS method utilizes feature extraction, selection, and machine learning integration to achieve precise and reliable estimation of cardiovascular signals. This enables real-time monitoring of physiological parameters with high accuracy.

The POS method is applicable in various domains such as healthcare, biometrics, human–computer interaction, and sports science. It can be used to monitor vital signs, detect stress, recognize emotions, and assess performance in various populations and environments without the need for invasive methods.

2.1.6. OMIT (Face2PPG)

The OMIT [8] method introduces a pioneering approach to rPPG, utilizing facial video data to extract cardiovascular signals, including heart rate and blood volume pulse, without a reliance on specialized sensors or hardware. Through the analysis of subtle colour variations in facial regions, OMIT provides a non-invasive and contactless method for physiological monitoring.

Central to the OMIT method are its key components, which begin with Facial Region of Interest detection. This stage involves the detection and segmentation of facial ROIs within video frames, typically achieved through facial landmark detection algorithms or deep learning-based techniques. These algorithms identify key facial landmarks, such as the eyes, nose, and mouth, acting as anchor points for ROI localization. Subsequently, colour signal extraction is performed, where the OMIT method extracts colour signals from these identified facial regions over time. By scrutinizing changes in pixel intensities across different wavelengths or colour channels, such as red, green, and blue, or chrominance channels, the method captures subtle variations in skin colour induced by cardiovascular activity.

Following colour signal extraction, temporal signal analysis is conducted to estimate cardiovascular parameters like heart rate and blood volume pulse. Techniques including Fourier analysis, autocorrelation, or peak detection algorithms are applied to identify periodic oscillations corresponding to physiological signals. Moreover, to enhance signal quality and robustness, the OMIT method incorporates motion compensation and artefact removal techniques. These techniques mitigate the impact of motion artefacts, facial expressions, and lighting variations on the extracted colour signals, ensuring accurate estimation of cardiovascular parameters.

The OMIT method offers several notable advantages. Firstly, it enables non-contact and non-invasive monitoring of cardiovascular signals, leveraging standard facial video data captured by off-the-shelf cameras or devices. This accessibility is further bolstered by the method's elimination of specialized sensors or hardware, enhancing convenience and accessibility for users across various settings. Additionally, OMIT facilitates real-time monitoring of cardiovascular activity, permitting continuous assessment of vital signs without interrupting daily activities or necessitating additional equipment.

The OMIT method has a wide range of applications in various fields such as healthcare, wellness monitoring, biometrics, and human–computer interaction. It can be used to instantly evaluate vital signs, detect stress, monitor fatigue, recognize emotions, and authenticate biometric data, among other uses.

2.2. Motion-Based Methods in RPPG

Motion-based methods in remote photoplethysmography leverage the analysis of motion information to extract cardiovascular signals, such as heart rate and blood volume pulse, from video data. By tracking subtle motion-induced changes in skin colour and brightness, motion-based methods offer a non-invasive and contactless approach to physiological monitoring. These methods entail several key components to facilitate signal extraction. Optical flow analysis involves tracking the apparent motion of pixels between consecutive frames of a video sequence, enabling the capture of subtle motion patterns on the skin surface induced by cardiovascular activity. Motion magnification techniques amplify these subtle motion signals in video data, enhancing their visibility for more accurate signal extraction. Additionally, frequency domain analysis, employing techniques like Fourier analysis or wavelet transform, extracts cardiovascular signals from motion signals by identifying periodic oscillations corresponding to heart rate or blood volume pulse in the frequency spectrum. Independent component analysis (ICA) further enhances signal fidelity by isolating motion-related components from other sources of noise or interference in video data. Motion-based methods offer advantages such as robustness to motion artefacts, dynamic signal extraction for real-time monitoring, and contactless monitoring, aligning with the broader goals of rPPG. These methods find applications in healthcare, fitness monitoring, human–computer interaction, and biometrics, enabling real-time assessment of vital signs, stress detection, activity tracking, emotion recognition, and biometric authentication.

2.2.1. The ICA Method

The ICA [9] method serves as a signal processing technique within remote photoplethysmography, aiming to disentangle mixed physiological signals from video data into statistically independent components. By decomposing video signals into these independent components, ICA facilitates the extraction of cardiovascular signals, such as heart rate and blood volume pulse, without direct skin contact. Fundamental to the ICA method is its signal decomposition phase, where mixed video signals are separated into independent components through the exploitation of statistical properties inherent in the data. Unlike conventional methods that are reliant on predetermined signal models, ICA identifies sources of variability within the data without prior assumptions, rendering it suitable for extracting complex physiological signals. Subsequently, upon identification of independent components, the ICA method proceeds with feature extraction, isolating relevant temporal patterns, frequency characteristics, or spatial distributions associated with cardiovascular activity within the video data. Following feature extraction, signal reconstruction is undertaken, wherein the original cardiovascular signals are regenerated from the independent components through the amalgamation of selected features and weighting coefficients, thus generating reconstructed signals closely resembling the underlying physiological activity captured by the video data. Additionally, the ICA method offers the capability of artefact removal, effectively eliminating unwanted artefacts or noise sources from the video data by segregating physiological signals from background noise or interference, thereby augmenting the fidelity and reliability of the extracted cardiovascular signals for subsequent analysis. The ICA method boasts several advantages, such as its capacity for unsupervised signal separation without prior knowledge of signal sources or characteristics, thereby enabling the extraction of complex physiological signals in diverse environments and conditions. Furthermore, its robustness to noise and artefacts present in the video data ensures accurate extraction of cardiovascular signals even amidst motion artefacts, lighting variations, or other sources of interference. Like other rPPG approaches, the ICA method affords non-contact and non-invasive monitoring of cardiovascular activity, obviating the necessity for physical sensors or devices attached to the body. The versatility of the ICA method spans across various domains including healthcare, wellness monitoring, biometrics, and human–computer interaction, where it finds application in the real-time assessment of vital signs, stress detection, emotion recognition, biometric authentication, and interactive systems.

2.2.2. Motion Magnification Method

The Motion Magnification method [10] constitutes a technique within rPPG utilized to amplify subtle motion signals embedded within video data, thereby enhancing the visibility of cardiovascular activity such as heart rate and blood volume pulse. By magnifying motion-induced colour and brightness variations on the skin surface, Motion Magnification facilitates non-invasive and contactless monitoring of physiological signals. Central to the Motion Magnification method are several key components essential for signal enhancement and extraction. Initially, motion detection algorithms are employed to discern subtle motion signals present in video frames, analyzing temporal variations in pixel intensities or colour information to pinpoint regions of interest (ROIs) manifesting motion-induced changes. Subsequently, motion amplification techniques are applied to boost these signals, thereby enhancing their visibility for subsequent signal extraction. Amplification methods may encompass spatial or temporal filtering, frequency domain analysis, or nonlinear transformations, selectively augmenting motion-related components while suppressing noise and artefacts. Following motion amplification, colour signals are extracted from the enhanced video data, representing variations in skin colour induced by changes in blood volume and flow modulated by cardiovascular activity. Finally, the Motion Magnification method undertakes signal processing and analysis of the extracted colour signals to estimate physiological parameters such as heart rate and blood volume pulse. Employing signal processing techniques like Fourier analysis, spectral analysis, or statistical modelling, characteristic oscillations corresponding to cardiovascular activity are identified, enabling accurate parameter estimation. The Motion Magnification method offers several advantages, including enhanced signal visibility, thereby facilitating easier extraction of cardiovascular signals from noisy or low-quality recordings. Furthermore, its robustness to motion artefacts and camera motion ensures accurate signal extraction in dynamic environments, while its non-invasive nature aligns with other rPPG approaches, allowing for contactless and non-invasive monitoring of cardiovascular activity without the need for physical sensors or devices attached to the body. With applications spanning across healthcare, wellness monitoring, sports science, and human–computer interaction domains, the Motion Magnification method finds utility in the real-time assessment of vital signs, stress detection, fatigue monitoring, emotion recognition, and performance evaluation.

2.2.3. Spatial Filtering Method

The Spatial Filtering method [11] represents a technique employed within rPPG to heighten the discernibility of cardiovascular signals by selectively filtering spatial components within video data. By mitigating noise and unwanted spatial frequencies while preserving pertinent physiological information, Spatial Filtering enhances the accuracy and dependability of cardiovascular signal extraction from video recordings.

The Spatial Filtering method encompasses several key components that are pivotal for signal enhancement and extraction. Initially, noise reduction techniques are implemented to quell unwanted spatial components in video data. This may entail employing spatial averaging, median filtering, or Gaussian smoothing to diminish high-frequency noise and artefacts while safeguarding signal integrity. Subsequently, following noise reduction, the Spatial Filtering method selects specific frequency bands corresponding to cardiovascular activity for further analysis. This selection process may involve bandpass filtering or Fourier analysis to isolate spatial frequencies linked to alterations in skin colour and brightness induced by blood flow. Following frequency band selection, the Spatial Filtering method elevates the visibility of cardiovascular signals by amplifying relevant spatial components while attenuating noise and interference. This enhancement procedure may incorporate adaptive filtering, Wiener filtering, or morphological operations to selectively bolster the signal-to-noise ratio in targeted frequency bands. Ultimately, the Spatial Filtering method scrutinizes temporal variations in the enhanced signals to estimate physiological parameters such as heart rate and blood volume pulse. Signal processing techniques like Fourier analysis, autocorrelation, or peak detection may be deployed to discern characteristic oscillations corresponding to cardiovascular activity.

The Spatial Filtering method provides multiple benefits, such as reducing noise to improve the quality and reliability of cardiovascular signal extraction. It also selectively enhances cardiovascular signals, improving their visibility while maintaining signal integrity and reducing false positives. In addition, similar to other remote photoplethysmography methods, Spatial Filtering allows for the non-contact and non-invasive monitoring of cardiovascular activity. This eliminates the requirement for physical sensors or devices attached to the body. The Spatial Filtering method is widely used in various fields such as healthcare, wellness monitoring, biometrics, and human–computer interaction. It is particularly useful for the real-time assessment of vital signs, stress detection, emotion recognition, biometric authentication, interactive systems, and other versatile applications.

2.2.4. Temporal Filtering Method

The temporal filtering method [11] constitutes a technique employed within remote photoplethysmography to amplify the visibility of cardiovascular signals by selectively filtering temporal components within video data. By eliminating noise and unwanted temporal frequencies while preserving pertinent physiological information, temporal filtering enhances the accuracy and reliability of cardiovascular signal extraction from video recordings.

The temporal filtering method encompasses several key components that are pivotal for signal enhancement and extraction. Initially, noise reduction techniques are applied to suppress unwanted temporal fluctuations in video data. This may involve employing temporal averaging, median filtering, or low-pass filtering to diminish high-frequency noise and artefacts while maintaining signal integrity. Subsequently, following noise reduction, the temporal filtering method selects specific frequency bands corresponding to cardiovascular activity for further analysis. This selection process may involve bandpass filtering or Fourier analysis to isolate temporal frequencies linked with alterations in skin colour and brightness induced by blood flow. Following frequency band selection, the temporal filtering method amplifies the visibility of cardiovascular signals by enhancing relevant temporal components while attenuating noise and interference. This enhancement procedure may incorporate adaptive filtering, spectral analysis, or wavelet decomposition to selectively bolster the signal-to-noise ratio in targeted frequency bands. Ultimately, the temporal filtering method analyzes enhanced signals to estimate physiological parameters such as heart rate and blood volume pulse.

Signal processing techniques such as autocorrelation, peak identification, and statistical modelling can be employed to detect distinct oscillations that correspond to cardiovascular activity. The temporal filtering approach has several advantages, including noise reduction to improve the quality and dependability of cardiovascular signal extraction. Additionally, it has the ability to specifically amplify cardiovascular signals, thus boosting their detectability without compromising signal accuracy and minimizing incorrect positive results. In addition, like other remote photoplethysmography techniques, temporal filtering enables the assessment of cardiovascular activity without the need for physical sensors or equipment connected to the body. The temporal filtering technique is extensively employed in diverse domains like healthcare, wellness monitoring, biometrics, and human–computer interaction. It is highly advantageous for the immediate evaluation of essential body functions, the identification of stress levels, the identification of emotions, verification of personal characteristics, and systems that allow for interaction.

2.2.5. Deep Learning Approach

Remote photoplethysmography [12] motion-based algorithms represent a cutting-edge advancement in physiological monitoring, harnessing the power of deep learning methodologies to revolutionize the extraction of cardiovascular signals from video data. At the core of these algorithms are convolutional neural networks (CNNs), which serve as foundational tools for spatial feature extraction from video frames. CNNs operate by employing multiple convolutional layers that hierarchically learn representations of image features, enabling the robust extraction of pertinent spatial information directly from raw pixel data. This spatial feature extraction process is pivotal in capturing subtle variations in skin colour and brightness induced by cardiovascular activity, laying the groundwork for accurate signal estimation.

Complementing the spatial analysis facilitated by CNNs, Recurrent Neural Networks (RNNs) play a crucial role in temporal modelling of physiological signals over time. Unlike traditional feedforward networks, RNNs possess the ability to capture sequential dependencies in data, making them ideally suited for analyzing dynamic cardiovascular signals within video sequences. By effectively encoding the temporal dynamics of physiological phenomena, RNNs contribute significantly to the accurate estimation of cardiovascular parameters such as heart rate and blood volume pulse.

Furthermore, autoencoders emerge as indispensable tools in the deep learning arsenal for rPPG motion-based algorithms. These unsupervised learning models excel in feature learning and dimensionality reduction, enabling the extraction of compact representations of video data. In the context of rPPG, autoencoders facilitate the identification of relevant physiological features that are essential for precise signal estimation, thereby enhancing the overall performance of the algorithm.

Moreover, the integration of Generative Adversarial Networks (GANs) and Transfer Learning techniques further enhances the robustness and generalization capabilities of deep learning models in rPPG. GANs are employed to generate synthetic data samples that closely resemble real physiological signals, thereby augmenting limited training data and improving the model's ability to adapt to diverse scenarios. Similarly, Transfer Learning leverages pre-trained deep learning models on large-scale image datasets to initialize model parameters for rPPG tasks, enabling efficient training on smaller datasets and enhancing model performance.

The advantages conferred by deep learning approaches in rPPG are manifold. Firstly, these methodologies enable end-to-end learning directly from raw video data, obviating the need for manual feature engineering and preprocessing, and thus streamlining the signal extraction process. Additionally, deep learning models exhibit robustness to variability in lighting conditions, skin tones, and facial expressions, owing to their capacity to learn complex and invariant representations of cardiovascular signals. This inherent robustness enhances the reliability and applicability of rPPG algorithms across diverse environments and populations. Furthermore, the scalability and adaptability of deep learning models render them versatile tools with wide-ranging applications in healthcare, wellness monitoring, biometrics, and human–computer interaction. From the real-time assessment of vital signs to stress detection, emotion recognition, biometric authentication, and interactive systems, deep learning approaches in rPPG offer unprecedented potential for advancing non-invasive physiological monitoring across various domains.

2.3. Multispectral Methods

Multispectral methods in rPPG [13] involve the analysis of photoplethysmographic signals across multiple spectral bands, typically beyond the visible spectrum, to extract cardiovascular information. These methods leverage the unique spectral absorption characteristics of hemoglobin to enhance signal robustness and accuracy in various environmental conditions and skin types.

Multispectral methods in rPPG utilize information from multiple spectral bands, including visible and near-infrared wavelengths, to extract cardiovascular signals from video data. By capturing physiological information across different spectral ranges, multispectral methods offer enhanced robustness to factors such as skin pigmentation, lighting variations, and motion artefacts, improving the accuracy and reliability of cardiovascular signal extraction.

The key components of multispectral methods are as follows:

- Wavelength Selection: Multispectral methods involve the selection of appropriate wavelengths that exhibit significant absorption variations due to changes in blood volume and oxygenation. These wavelengths may span visible, near-infrared, and sometimes infrared ranges to capture hemoglobin absorption features while minimizing interference from other sources, such as melanin or ambient light.
- Spectral Decomposition: Once video data are acquired across multiple spectral bands, multispectral methods decompose the signals to extract spectral components related to cardiovascular activity. This may involve spectral analysis techniques such as principal component analysis (PCA) or independent component analysis (ICA) to identify spectral features associated with hemoglobin absorption changes.
- Signal Fusion: After spectral decomposition, multispectral methods integrate information from different spectral bands to enhance the visibility of cardiovascular signals. Signal fusion techniques, such as weighted averaging or spectral combination, merge spectral components to improve signal-to-noise ratio and signal fidelity, resulting in more robust cardiovascular signal extraction.
- Temporal Signal Analysis: Finally, multispectral methods analyze temporal variations in the fused signals to estimate physiological parameters such as heart rate and blood volume pulse. Signal processing techniques such as Fourier analysis, wavelet transform, or machine learning may be employed to extract and analyze temporal features from multispectral data.

Robustness to Skin Pigmentation: Multispectral methods are less sensitive to variations in skin pigmentation compared to single-band approaches, as they utilize information from multiple spectral ranges to extract cardiovascular signals.

Enhanced Signal Fidelity: By capturing physiological information across different spectral bands, multispectral methods offer enhanced signal fidelity and robustness to noise, motion artefacts, and lighting variations, improving the accuracy of cardiovascular signal extraction.

Versatility in Environmental Conditions: Multispectral methods are adaptable to diverse environmental conditions and lighting scenarios, making them suitable for use in various applications and settings, including indoor and outdoor environments.

Multispectral methods have applications in healthcare, wellness monitoring, biometrics, and human–computer interaction. They can be utilized for the real-time assessment of vital signs, stress detection, emotion recognition, biometric authentication, and interactive systems.

2.3.1. Near-Infrared Videos

NIR Imaging in rPPG [13] involves the utilization of near-infrared light to capture physiological signals, such as heart rate and blood volume pulse, from video data. Near-infrared light penetrates deeper into the skin compared to visible light, allowing for the measurement of hemoglobin absorption changes and providing valuable information about cardiovascular activity.

At the heart of NIR Imaging lie several key components that are essential for effective signal acquisition and processing. Firstly, NIR Imaging setups employ light sources emitting near-infrared wavelengths typically ranging from 700 to 1000 nanometres. These light sources, often light-emitting diodes or laser diodes, offer narrow spectral bandwidth and controllable intensity, ensuring precise illumination for optimal signal capture. Paired with specialized cameras equipped with sensors sensitive to near-infrared light, NIR Imaging systems are designed to capture subtle hemoglobin absorption changes. These cameras may be purpose-built for NIR imaging or modified with optical filters to selectively block out visible light and capture NIR wavelengths, thereby enhancing signal fidelity.

Furthermore, the integration of optical filters, particularly bandpass filters, is crucial in isolating NIR wavelengths from ambient light and other sources of interference. By selectively transmitting near-infrared light while attenuating unwanted spectral components, optical filters improve the signal-to-noise ratio, thereby enhancing the visibility of cardiovascular signals. Subsequent signal processing techniques, such as spatial and temporal filtering, frequency domain analysis, and machine learning algorithms, are applied to extract and analyze hemoglobin absorption variations associated with cardiovascular activity from the captured NIR video data.

The advantages offered by NIR Imaging are multifaceted. Firstly, near-infrared light's deep tissue penetration capability enables the measurement of hemoglobin ab-sorption changes in deeper tissue layers, enhancing the effectiveness of cardiovascular signal capture, particularly in regions with higher blood perfusion. Additionally, NIR Imaging exhibits robustness to lighting conditions and skin pigmentation, as near-infrared light can penetrate through skin and ambient light, ensuring consistent and reliable signal acquisition across various environments. Moreover, NIR Imaging facilitates contactless and non-invasive monitoring of cardiovascular signals, aligning with the core principles of rPPG and eliminating the need for physical sensors or devices attached to the body.

The applications of NIR Imaging span diverse domains, including healthcare, wellness monitoring, biometrics, and human–computer interaction. From real-time assessment of vital signs to stress detection, emotion recognition, biometric authentication, and interactive systems, NIR Imaging holds promise for advancing non-invasive physiological monitoring in various contexts.

In low-light or special-environment cases, it is impossible to supply the light needed. For such measurements it is suggested to use near-infrared domain. The utilization of the near-infrared spectrum for the estimation of heart rate (HR) presents a distinct advantage. This is primarily due to the fact that it can intensify the light without negatively impacting human perception, owing to the use of an invisible NIR light source. In low-light or dark conditions, the estimation of heart rate can be achieved while minimizing the impact of ambient light fluctuations that are more prevalent in the visible domain. This is because visible wavelengths are the only light emitted by common sources of artificial light, like monitors or television screens [13].

A proposed algorithm using this method is show in Figure 1 [13].

The objective in this instance is to estimate HR for a specified time period (30 s in described experiments), assuming that HR does not fluctuate significantly throughout that time. In this paper, there is a unique temporal fusion method provided to enhance the resilience against head movements that could occur inside the time span. While it is possible that significant head movement will occur over the designated time period, it was expected that in practical HR monitoring situations, such as while driving or in an office, there would be brief intervals characterized by minimal head movement. As a result, the time period is divided into brief sub-windows via the sliding window method (five-second recordings with one-second intervals [13]). Following this, a candidate HR histogram was constructed for each sub-window utilizing the spatial and spectral face patch sampling-based HR estimation. The histogram that was produced as a whole was fused to form the final histogram. The final HR estimate is determined by majority vote and parabola fitting, as seen on the right side of Figure 1. This method is predicated on the assumption that the candidate HRs can be accurately and consistently measured from the brief time intervals that are less impacted by head movements.

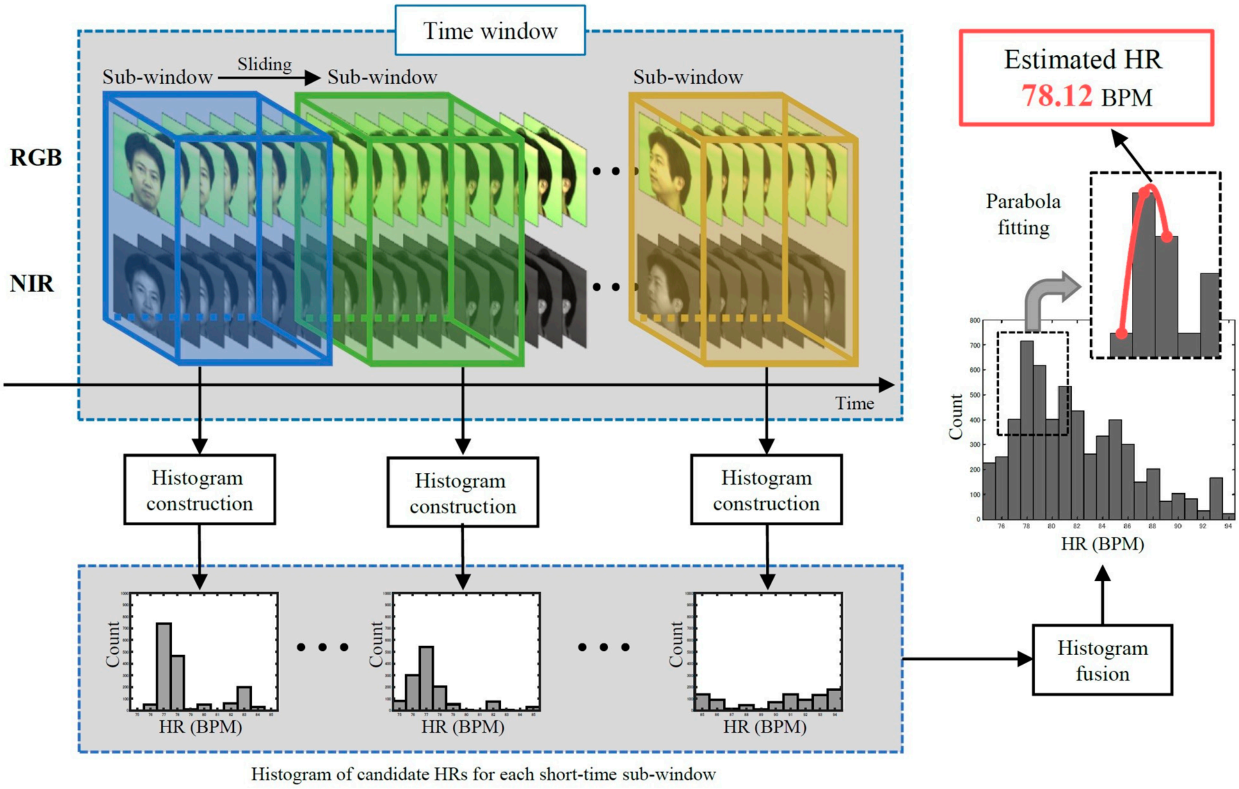

Figure 1. Method that estimates heart rate using an RGB-NIR face video as an overall framework.

In this paper, we also describe a face tracking algorithm that minimizes noises as body movement influence. To obtain the final HR estimate for the examined time window, the produced histograms for each short-time sub-window are fused to form the final histogram. In the final histogram, the most voted HR bin is deemed the most reliably and consistently computed HR utilizing stable video sub-regions less impacted by light fluctuations and head movements in the spatial–spectral–temporal domain. Parabola fitting is used to acquire the final HR estimate in the real value precision by utilizing that bin along with its nearby bins. The estimated HR for the examined time frame is shown by the apex of the fitted parabola.

The NIR method with comparison to the RGB method was proven to be more accurate, required less processing, and was faster.

2.3.2. Hemoglobin Spectroscopy

Hemoglobin spectroscopy in rPPG involves the analysis of the spectral absorption characteristics of hemoglobin to extract cardiovascular signals from video data. By measuring changes in hemoglobin concentration and oxygenation levels, hemoglobin spectroscopy provides valuable insights into cardiovascular activity and enables non-invasive monitoring of physiological parameters.

Central to hemoglobin spectroscopy setups are several key components that are essential for effective signal acquisition and analysis. Firstly, light sources emitting a range of wavelengths across the visible and near-infrared spectrum are utilized. These light sources, which may include LEDs, laser diodes, or broadband light sources with adjustable spectral characteristics, ensure precise illumination for optimal signal capture. Following skin illumination with light of varying wavelengths, spectral analysis techniques are employed to measure hemoglobin absorption. Reflectance spectroscopy, transmittance spectroscopy, or diffuse optical spectroscopy methods are commonly utilized to analyze spectral absorption features associated with hemoglobin.

In some hemoglobin spectroscopy setups, multispectral imaging techniques are employed to capture spatially resolved spectral data from video frames. Multispectral cameras equipped with sensors sensitive to different wavelengths enable the acquisition of spectral information across multiple spatial locations simultaneously, facilitating the analysis of hemoglobin absorption variations across the skin surface. Subsequently, signal processing techniques are applied to extract cardiovascular signals from the measured hemoglobin absorption spectra. Mathematical modelling, spectral decomposition, or machine learning algorithms may be employed for this purpose, enabling the identification and analysis of spectral features associated with cardiovascular activity.

The advantages offered by hemoglobin spectroscopy are manifold. Firstly, it provides direct measurements of hemoglobin absorption characteristics, offering insights into changes in blood volume and oxygenation levels associated with cardiovascular activity. Additionally, hemoglobin spectroscopy offers high sensitivity and specificity for cardiovascular signal extraction by analyzing spectral the absorption features of hemoglobin. This method's versatility in applications spans across healthcare, wellness monitoring, sports science, and human–computer interaction domains, offering versatile and non-invasive methods for physiological monitoring.

2.3.3. Dual-Wavelength Imaging

Dual-Wavelength Imaging in rPPG involves the simultaneous capture of video data at two distinct wavelengths of light, typically in the visible and near-infrared spectrum. By measuring changes in light absorption at these wavelengths, Dual-Wavelength Imaging enables the extraction of cardiovascular signals, providing valuable insights into blood volume and oxygenation changes in the skin.

At the core of Dual-Wavelength Imaging setups lie several key components essential for effective signal acquisition and analysis. Firstly, light sources emitting at two distinct wavelengths, commonly in the visible and near-infrared spectrum, are employed. LEDs or laser diodes are often utilized for this purpose, with one emitting visible light (e.g., green or red) and the other emitting near-infrared light. Following skin illumination with dual wavelengths of light, spectral separation techniques are employed to isolate the signals captured at each wavelength. Optical filters or dichroic mirrors facilitate the separation of the reflected or transmitted light into distinct spectral bands for further analysis.

Subsequently, signal processing techniques are applied to extract cardiovascular signals from the captured images. This involves analyzing changes in light intensity or colour at each wavelength to estimate parameters such as heart rate and blood volume pulse. To ensure accurate signal extraction, Dual-Wavelength Imaging setups often synchronize the acquisition of video data at both wavelengths. This ensures that changes in blood volume and oxygenation are captured simultaneously, facilitating robust cardiovascular signal extraction.

The benefits provided by Dual-Wavelength Imaging are remarkable. Firstly, it allows for the assessment of hemoglobin absorption at two different wavelengths, which increases the sensitivity to changes in blood volume and oxygenation levels. In addition, Dual-Wavelength Imaging provides increased resistance to motion artefacts and environmental influences by recording images at two wavelengths simultaneously, boosting the precision and dependability of cardiovascular signal extraction. The versatility of this technology extends to other sectors, such as healthcare, wellness monitoring, sports science, and human–computer interaction. It provides flexible and non-invasive techniques for measuring physiological parameters.

2.4. Depth-Based Methods

Depth-based methods in rPPG utilize depth sensing technologies, such as structured light or time-of-flight cameras, to capture three-dimensional (3D) facial geometry and extract cardiovascular signals from video data. By analyzing changes in blood perfusion across facial surfaces, depth-based methods offer a non-contact and robust approach to physiological monitoring.

The key components of depth-based methods are as follows:

- Depth Sensing Technology: Depth-based methods utilize depth sensing technologies to capture 3D facial geometry, typically through structured light projection or time-of-flight measurements. These technologies enable the accurate reconstruction of facial surfaces and provide depth information for each pixel in the captured images.
- Surface Reconstruction: After capturing depth data, depth-based methods reconstruct facial surfaces from the depth images. Surface reconstruction techniques, such as triangulation or depth map fusion, are employed to generate high-resolution 3D models of facial geometry, which provide spatial information about the skin surface.
- Blood Perfusion Analysis: Once facial surfaces are reconstructed, depth-based methods analyze changes in blood perfusion across the skin surface to extract cardiovascular signals. This may involve measuring variations in skin colour or intensity, which are indicative of changes in blood volume and oxygenation levels induced by cardiovascular activity.
- Signal Processing and Analysis: Following blood perfusion analysis, signal processing techniques are applied to extract cardiovascular signals from the depth data. This may involve spatial and temporal filtering, frequency domain analysis, or machine learning algorithms to identify and analyze physiological features associated with cardiovascular activity.

Non-Contact Monitoring: Depth-based methods allow for the monitoring of cardiovascular signals without the need for physical sensors or devices that are attached to the body. This approach is non-invasive and does not require any direct contact.

Depth-based approaches exhibit a higher level of resilience towards environmental influences, such as changes in lighting, motion disturbances, and variations in skin pigmentation, when compared to conventional imaging techniques. This characteristic makes them well suited for application in diverse contexts and under various conditions.

Spatially Dense Measurements: Depth-based methods offer precise and accurate measurements of facial surfaces, allowing for thorough investigation of blood perfusion patterns in various parts of the face. This leads to enhanced accuracy in extracting cardiovascular signals.

Depth-based methods are utilized in several fields, such as healthcare, wellness monitoring, biometrics, and human–computer interaction. They can be employed for the immediate evaluation of essential indications, the identification of stress, the recognition of emotions, biometric verification, and interactive systems, among various other uses.

2.4.1. Depth PPG

Depth-based methods [14] in rPPG utilize depth sensing technologies, such as structured light or time-of-flight cameras, to capture three-dimensional facial geometry and extract cardiovascular signals from video data. Unlike conventional imaging methods that primarily rely on colour variations in 2D images, Depth PPG harnesses depth information to discern subtle changes in blood perfusion across the skin surface, thereby offering heightened accuracy and robustness in physiological assessment.

At the heart of Depth PPG lies several key components that are pivotal for its efficacy in capturing and analyzing cardiovascular signals. Depth sensing technologies, such as structured light or time-of-flight cameras, serve as the cornerstone of Depth PPG systems, enabling the capture of 3D facial geometry with exceptional spatial resolution and accuracy.

These sensors emit structured light patterns or measure the time-of-flight of light to generate detailed depth maps, furnishing comprehensive depth information for each pixel in the acquired images.

Subsequent to depth data acquisition, Depth PPG engages in temporal analysis to discern cardiovascular signals by scrutinizing minute variations in facial geometry over time. This meticulous tracking facilitates the detection of subtle alterations in blood perfusion induced by cardiovascular activity. The estimation of blood perfusion across the skin surface ensues, facilitated by the analysis of changes in depth data. As blood traverses facial blood vessels, it modulates the optical properties of the skin, thereby eliciting slight displacements in facial tissue discernible through depth sensing technology.

Following blood perfusion estimation, Depth PPG employs signal processing techniques to extract cardiovascular signals from the depth data. This entails spatial and temporal filtering, frequency domain analysis, or the application of machine learning algorithms to identify and analyze physiological features associated with cardiovascular activity.

Depth PPG proffers several advantages over traditional imaging methods. Its direct measurement of changes in blood perfusion across the skin surface contributes to enhanced accuracy in cardiovascular signal extraction, circumventing the reliance solely on colour variations in 2D images. Moreover, Depth PPG exhibits robustness to environmental factors such as lighting variations, motion artefacts, and skin pigmentation, rendering it suitable for application in diverse settings and conditions. Like other rPPG modalities, Depth PPG facilitates non-contact and non-invasive monitoring of cardiovascular signals, obviating the necessity for physical sensors or devices affixed to the body.

With its multifaceted advantages, Depth PPG finds applications across various domains including healthcare, wellness monitoring, biometrics, and human–computer interaction. Its utility extends to the real-time assessment of vital signs, stress detection, emotion recognition, biometric authentication, and interactive systems, underscoring its versatility and significance in advancing physiological monitoring paradigms.

2.4.2. Three-Dimensional Convolutional Networks

Three-Dimensional Convolutional Networks (3D CNNs) [15] in rPPG are deep learning architectures specifically designed to analyze temporal and spatial features in three-dimensional video data captured from depth sensors or multi-view camera setups. By leveraging the spatial and temporal information encoded in 3D video sequences, 3D CNNs enable robust extraction of cardiovascular signals, offering enhanced accuracy and reliability in physiological monitoring.

Central to the architecture of 3D Convolutional Networks are multiple layers of 3D convolutional, pooling, and fully connected layers. These networks possess the capacity to learn hierarchical representations of spatiotemporal features directly from 3D video data, thereby enabling end-to-end extraction of cardiovascular signals without the necessity for manual feature engineering. The input to 3D CNNs comprises 3D video sequences obtained from depth sensors or multi-view camera setups, with each frame in the sequence containing depth information. This configuration allows the network to scrutinize spatial features across various facial regions over time.

Temporal analysis in 3D CNNs is facilitated through the convolution of 3D kernels across the temporal dimension of the input video sequences. This mechanism empowers the network to apprehend temporal dynamics and discern subtle changes in cardiovascular signals over time, such as variations in blood perfusion induced by heartbeats. Concurrently, spatial analysis is conducted by analyzing spatial features across different facial regions within each frame of the input video sequences. By convolving 3D kernels across the spatial dimensions, the network can capture spatial patterns associated with blood perfusion and extract cardiovascular signals from specific facial regions.

The effectiveness of 3D CNNs lies in their ability to learn spatial and temporal representations, allowing for accurate extraction of cardiovascular signals without the need for manual feature engineering. Moreover, 3D convolutional neural networks enable the extraction of signals directly from unprocessed 3D video data, eliminating the necessity for preprocessing or manual extraction of features. These networks are particularly resistant to changes in facial appearance, lighting conditions, and facial expressions, making them well suited for use in a wide range of situations and circumstances.

3D Convolutional Networks have numerous advantages and are used in various sectors such as healthcare, wellness monitoring, biometrics, and human–computer interaction. Their ability to be used in the real-time assessment of vital signs, stress detection, emotion recognition, biometric authentication, and interactive systems highlights their importance in developing physiological monitoring methods.

2.5. Hybrid or Combined Models

Hybrid Models in rPPG combine multiple approaches, such as colour-based methods, motion-based methods, Spatial Filtering, temporal filtering, deep learning techniques, and others, to enhance the accuracy and robustness of cardiovascular signal extraction from video data. By leveraging the complementary strengths of different methods, Hybrid Models offer improved performance and versatility in physiological monitoring.

At the core of Hybrid Models lies the integration of multiple rPPG approaches into a cohesive framework. This integration encompasses diverse methodologies such as colour-based methods, motion-based techniques, spatial and temporal filtering, and deep learning frameworks. By amalgamating these methodologies, Hybrid Models capitalize on varied information sources and exploit the complementary strengths of each approach to bolster signal extraction.

Feature fusion constitutes a pivotal aspect of Hybrid Models, where features extracted from different methodologies are combined into a unified representation. This fusion process may entail concatenating feature vectors, applying feature transformation techniques, or leveraging machine learning algorithms to learn optimal feature representations. Through feature fusion, Hybrid Models effectively integrate information from diverse methodologies, enhancing the richness and robustness of the extracted signals.

Some Hybrid Models leverage model ensemble techniques to further enhance performance and robustness. Model ensemble methods such as bagging, boosting, or stacking enable the aggregation of predictions from multiple individual models trained on different rPPG approaches. By combining predictions from diverse models, ensemble techniques bolster overall performance and robustness, thereby elevating the efficacy of Hybrid Models in physiological monitoring.

Adaptive fusion strategies represent another key component of Hybrid Models, where the contribution of each methodology is dynamically adjusted based on the quality and reliability of the extracted signals. These adaptive fusion techniques ensure optimal utilization of information from different approaches and enhance the adaptability of the model to changing conditions, thereby bolstering its efficacy in diverse environments and scenarios.

Implementing Hybrid Models results in concrete advantages in the field of physiological monitoring. Hybrid Models use the strengths of various rPPG techniques to achieve improved performance and reliability in extracting cardiovascular signals. Due to their versatility, they may be adjusted to different ambient circumstances, lighting situations, and subject attributes. This makes them useful for a wide range of applications in healthcare, wellness monitoring, biometrics, and human–computer interaction. Hybrid Models represent the forefront of physiological monitoring paradigms, since they possess the ability to check vital signs in real time, identify stress, recognize emotions, authenticate biometrics, and interact with users.

3. Tools and Datasets

As for research, there is rPPG-Toolbox, which is the most used tool for rPPG signal processing. It is an open-source platform designed for camera-based physiological sensing, also known as remote photoplethysmography (rPPG). The toolbox currently supports seven datasets, namely SCAMPS: Synthetics for Camera Measurement of Physiological Signals [16], UBFC-rPPG (which stands for Univ. Bourgogne Franche-Comté Remote Photoplethysmography) [17], PURE [18], BP4D+ [19], UBFC-Phys [20], MMPD (Multi-domain Mobile Video Physiology Dataset) [21], and iBVP [22].

The rPPG-Toolbox not only evaluates the current cutting-edge neural and unsupervised techniques, but it also facilitates the adaptable and swift creation of your own algorithms. Figure 2 shows an overview of its capabilities.

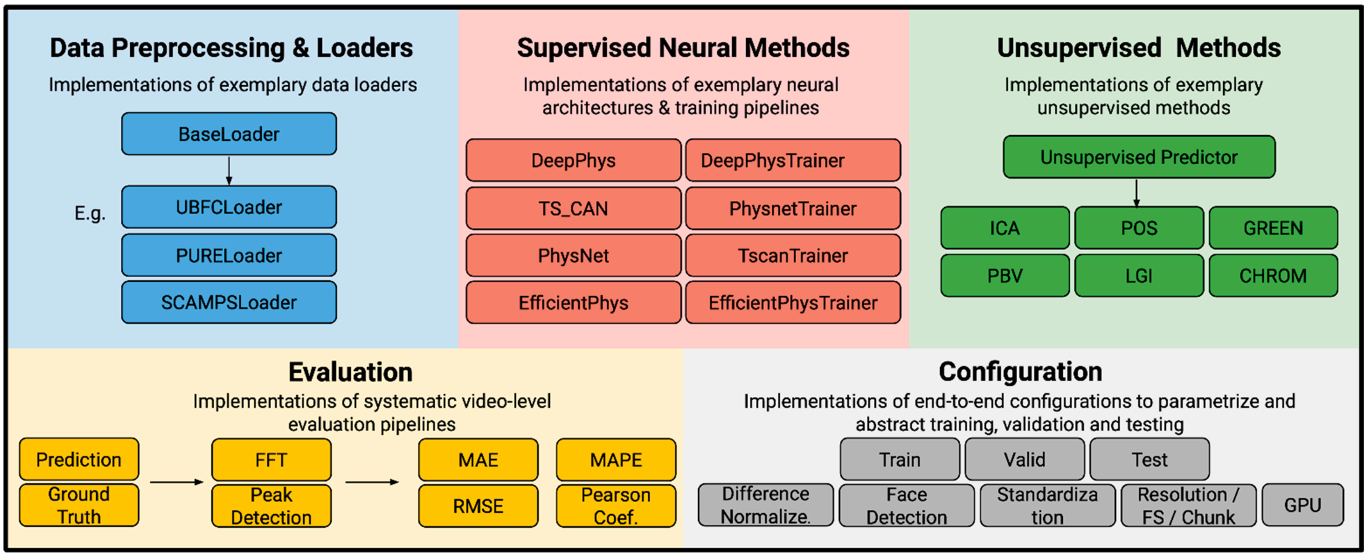

Figure 2. An overview of the rPPG-Toolbox functions.

The utilization of cameras and computational algorithms for the non-invasive, cost-effective, and easily expandable assessment of physiological vital signs, such as cardiac and pulmonary functions, is highly appealing. Obtaining diverse data that encompass various environments, body motions, illumination conditions, and physiological states is a challenging and costly process that requires significant time and effort. Data are crucial in both the training and evaluation of machine learning models. Nevertheless, the process of generalization may be compromised if the training data lack representativeness, and the evaluation of the model's performance can be difficult if the testing data do not encompass the required variations and diversity. The utilization of public datasets has made a substantial contribution to the comprehension of algorithmic performance in this particular field. Collecting these datasets requires a significant amount of time and effort. They also contain highly personal and sensitive biometric information, such as facial videos and physiological waveforms, which can be used to identify individuals. Obtaining datasets that encompass a diverse range of examples across various cardiac and pulmonary parameters, such as heart and breathing rates and variabilities, pulse arrival times, and waveform morphologies, poses a challenge. In addition, the majority of these datasets are gathered in one specific place, resulting in a lack of diversity in terms of subject appearance, lighting conditions, context, and behaviours. Table 1 provides an overview of the characteristics of these datasets, including their accessibility to researchers in both industry and academia, where "+" sign means available, and "-" not available.

Table 1. Summary of camera physiological measurement datasets.

Dataset Name	Number of Subjects in Dataset	Number of Videos in Dataset	Standard Data Provided in Dataset	Subject Diversity	Environment Diversity	Free Access
MAHNOB	27	527	Electrocardiogram waveform, Electroencephalogram waveforms, Breathing waveform.	-	-	+
BP4D+ [2]	140	1400	Blood pressure waveform, Action Units.	+	-	-
VIPL-HR	107	3130	Photoplethysmogram waveform, HR, Blood oxygenation.	-	-	+
COHFACE	40	160	Photoplethysmogram waveform.	-	-	+
UBFC-RPPG [3]	42	42	Photoplethysmogram waveform, Pulse rate.	-	-	+
UBFC-PHYS [4]	56	168	Photoplethysmogram waveform, Electrodermal activity.	+	-	+
RICE CamHRV	12	60	Photoplethysmogram waveform.	-	-	+
MR-NIRP	18	37	Photoplethysmogram waveform.	-	-	+
MMPD [5]	33	660	Photoplethysmogram waveform.	+	+	+
PURE [6]	10	59	Photoplethysmogram waveform, Blood oxygenation.	-	-	+
iBVP [7]	32	32	Breathing waveform, Photoplethysmogram waveform.	-	-	+
rPPG	8	52	Pulse rate, Blood oxygenation.	-	-	+
OBF	106	212	Photoplethysmogram waveform, Electrocardiogram waveform, Blood pressure waveform.	-	-	-
PFF	13	85	Pulse rate.	-	-	+
VicarPPG	20	10	Photoplethysmogram waveform.	-	-	+
CMU	140	140	Pulse rate.	+	+	+
SCAMPS [8]—synthetic dataset	2800	2800	Photoplethysmogram waveform, Pulse rate, breathing waveform, Breathing rate, Action units.	+	+	+

From the available datasets, the most rich in subjects and environment diversity is SCAMPS. It is useful in various domains of machine learning, but it is not readily accessible for the purpose of measuring physiological states using cameras. Synthetic data provide labels that are free from noise and have precise synchronization. These labels may be unattainable through other means, such as precise pixel-level segmentation maps. Additionally, synthetic data allow for a high level of control over the variation and diversity present in the dataset. SCAMPS [16] is a collection of synthetic data consisting of 2800 videos (equivalent to 1.68 million frames) that include synchronized cardiac and respiratory signals as well as facial action intensities. The RGB frames are accompanied by segmentation maps.

SCAMPS offers accurate descriptive statistics regarding the fundamental waveforms, such as inter-beat interval, heart rate variability, and pulse arrival time. Ultimately, it showcases the initial outcomes achieved by training on these artificially generated data and can be used for evaluation on real-world datasets to demonstrate the ability to apply the learned knowledge to different scenarios. Figure 3 shows the workflow of creating synthetic videos.

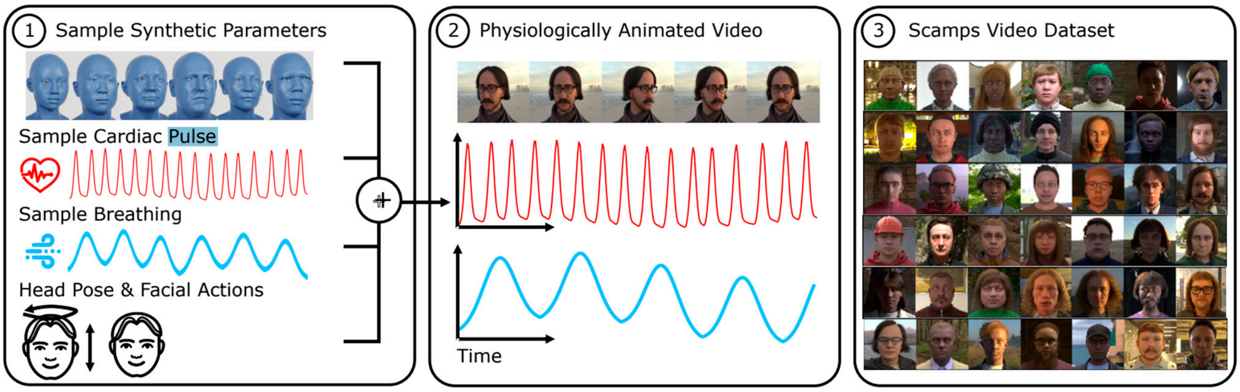

Figure 3. A dataset of synthetic videos with aligned physiological and behaviour signals [16].

The BP4D+ [23] dataset is second largest available dataset by subject variability and video feed-wise. It is composed of a Multimodal Spontaneous Emotion Corpus (MMSE), which contains multimodal datasets including synchronized 3D, 2D, thermal, and physiological data sequences (e.g., heart rate, blood pressure, skin conductance (EDA), and respiration rate), as well as meta-data (facial features and FACS codes) (Figure 4). There are 140 subjects, including 58 males and 82 females, with ages ranging from 18 to 66 years old. Ethnic/racial ancestries include Black, White, Asian (including East Asian and Middle Eastern), Hispanic/Latino, and others (e.g., Native American). With 140 subjects and 10 tasks (emotions) for each subject included in the database, there are over 10TB high-quality data points generated for the research community.

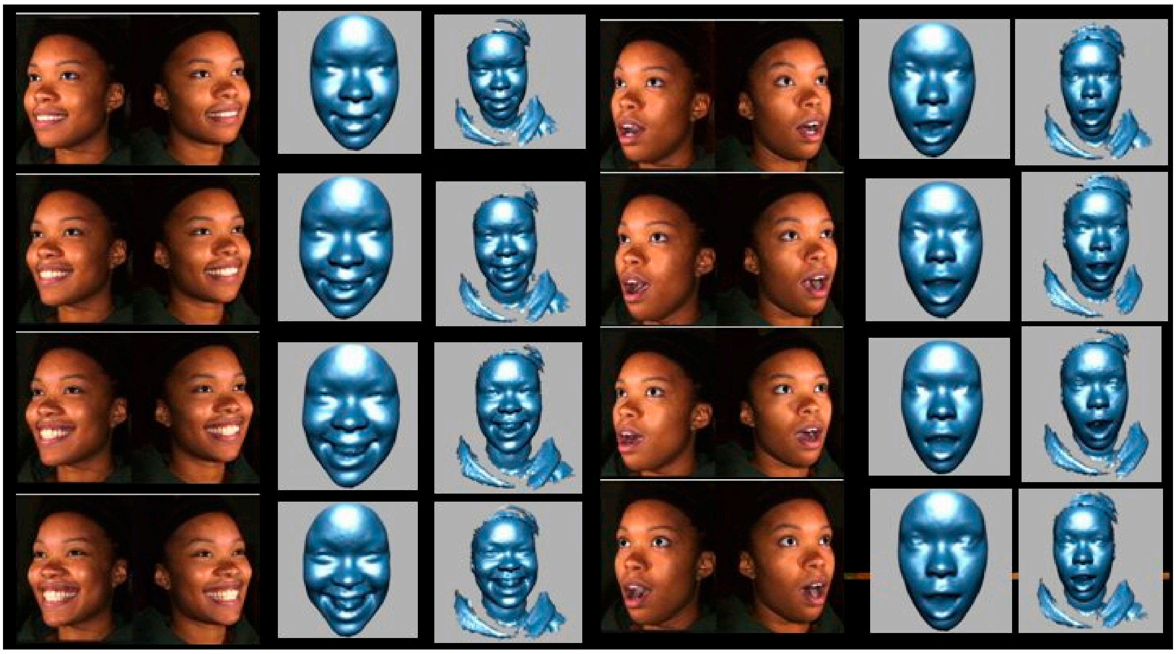

Figure 4. BP4D+ dataset representation and generation.

UBFC-rPPG [17] comprises two datasets which are focused specifically on rPPG analysis. The UBFC-RPPG [17] database was established utilizing a bespoke C++ programme for video capture employing an affordable webcam (Logitech C920 HD Pro) operating at a frame rate of 30 frames per second and a resolution of 640 × 480. The video was stored in an uncompressed 8-bit RGB format. The ground-truth PPG data, which include the PPG waveform and the PPG heart rates, were obtained using a Contec Medical CMS50E transmissive pulse oximeter. During the recording, the subject assumed a seated position directly facing the camera, positioned approximately 1 m away, ensuring their face was clearly visible. All experiments were carried out indoors, with different levels of sunlight and indoor lighting.

PURE (Pulse Rate Detection Dataset) [24] consists of 10 persons performing different, controlled head motions in front of a camera. Throughout these sentences, the image sequences of the head and reference pulse measurements were captured. There were 10 individuals, consisting of 8 males and 2 females, who were observed in six distinct scenarios, resulting in a total of 60 one-minute sequences. The videos were recorded using an eco274CVGE camera manufactured by SVS-Vistek GmbH, Gilching, Germany. The frame rate of the videos was 30 Hz, and they were captured at a cropped resolution of 640 × 480 pixels. The camera was equipped with a 4.8 mm lens. Data were collected simultaneously using a finger clip pulse oximeter (pulox CMS50E) that provides readings of pulse rate wave and SpO2 with a sampling rate of 60 Hz. The test subjects were positioned in front of the camera at an average distance of 1.1 m. The lighting condition was natural daylight coming through a large window directly in front of the face, with clouds causing slight variations in illumination over time. The test subjects were positioned in front of the camera at an average distance of 1.1 m.

The UBFC-Phys [20] dataset is a publicly available multimodal dataset. It consists of data from 56 participants who took part in an experiment following a strict protocol based on the widely recognized Trier Social Stress Test (TSST) (Figure 5). The experiment was conducted in three stages: a period of rest, a speech task, and arithmetic tasks of varying difficulty levels.

Figure 5. Samples of UBFC-Phys dataset recordings.

Throughout the experiment, the participants were recorded on camera and equipped with a wristband that monitored their blood volume pulse (BVP) and electrodermal activity (EDA) signals. Prior to commencing the experiment and upon its conclusion, participants filled out a questionnaire to determine their self-reported anxiety scores. The dataset includes video recordings of participants and BVP/EDA signals recorded during the three tasks, along with their anxiety scores measured before and after the experimental sessions.

MMPD (Multi-domain Mobile Video Physiology Dataset) [21] consists of 11 h (1152 K frames) of recordings captured from mobile phones, involving 33 subjects. The dataset was specifically created to encompass a wider range of skin tones, body movements, and lighting conditions in the captured videos.

The iBVP [22] dataset consists of synchronized RGB and thermal infrared videos, along with PPG ground-truth signals, obtained from an ear. The PPG signals are labelled with manual signal quality indicators, as well as with the SQA-PhysMD model that was trained and validated to perform detailed (per-sample) assessment of signal quality. The data acquisition protocol was designed to elicit real-world fluctuations in psycho-physiological states, as well as head movement. Each participant underwent four conditions, namely rhythmic slow breathing and rest, an easy math task, a difficult math task, and a guided head movement task. RGB and thermal cameras were placed in front of the participant at a distance of approximately 1 m. A Logitech BRIO 4K UHD webcam was utilized to record RGB video frames with a resolution of 640 × 480. Simultaneously, an FLIR system's A65SC thermal camera (Wilsonville, OR, USA)was employed to capture thermal infrared frames with a resolution of 640 × 512. The frame rate for both RGB and thermal acquisition was set to 30 frames per second (FPS). The dataset consists of 124 sessions, each lasting 3 min, resulting in a total of 372 min (approximately 6 h) of RGB-Thermal video recordings.

4. Tests

To process the dataset, Visual Studio Code (VSCode v1.85) from Microsoft® was used as programming tool. rPPG-Toolbox source code was cloned from GitHub and set up according to its instructions [25]. There was a defined Python(v3.8) version and environment to run the scripts and process datasets. Toolbox allows to process all the datasets pointed in configuration file, using all the supported methods at once, but as there was limited storage capacity the laptop they were executed on, the test was divided into separate, smaller tasks as shown in Figure 6.

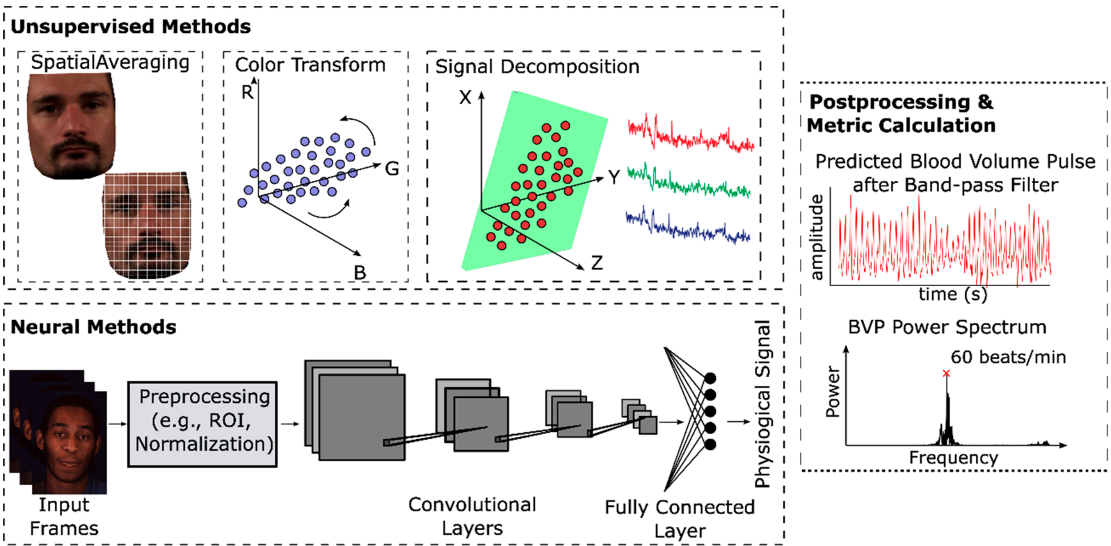

Figure 6. Flow of rPPG-Toolbox dataset pipeline from processing, training, interface, and evaluation [25].

There is also the possibility of looking through and visualizing the process of processing video files with benchmark trends and values. Figure 7 shows an original UBFC-Phys dataset video, Figure 8 shows the processing of the video, and Figure 9 shows the waveforms.

Figure 7. Original UBFC-Phys dataset video.

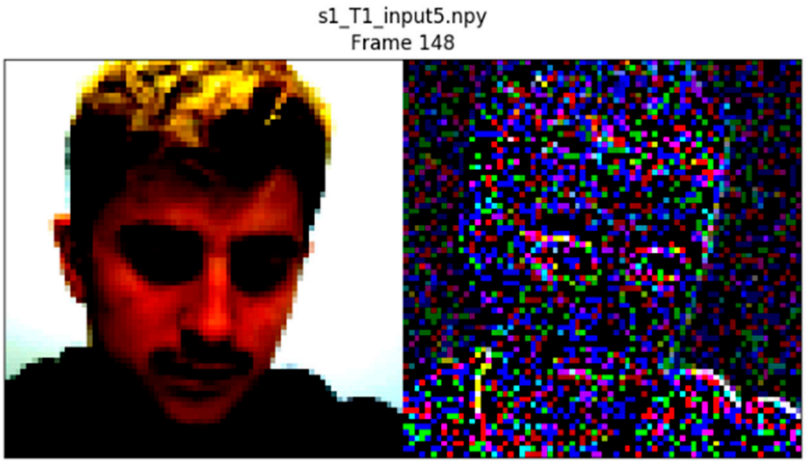

Figure 8. Processed data in Python notebook.

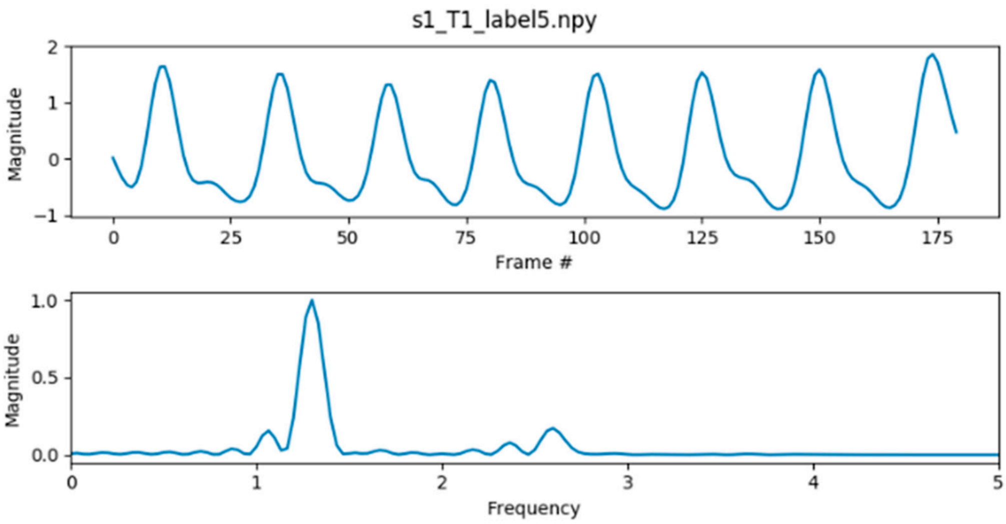

Figure 9. Waveforms of magnitude in frame.

5. Results

Comparing different unsupervised and supervised methods for remote photoplethysmography on multiple datasets shows significant differences in performance, providing important insights into their effectiveness, reliability, and applicability. The results of the test are represented in Table 2.

Table 2. rPPG-Toolbox results of run datasets.

			Test Set							
			PURE [9]		**UBFC-rPPG [3]**		**UBFC-Phys [4]**		**MMPD [5]**	
	Method	**Train Set**	**MAE ↓**	**MAPE ↓**	**MAE ↓**	**MAPE ↓**	**MAE ↓**	**MAPE ↓**	**MAE ↓**	**MAPE ↓**
UNSUPERVISED	GREEN [10]	N/A	10.09	10.28	19.81	18.78	13.55	16.01	21.68	24.39
	ICA [11]	N/A	4.77	4.47	14.70	14.34	10.03	11.85	18.60	20.88
	CHROM [12]	N/A	5.77	11.52	3.98	3.78	4.49	6.00	13.66	15.99
	LGI [13]	N/A	4.61	4.96	15.80	14.70	6.27	7.83	17.08	18.98
	PBV [14]	N/A	3.91	4.82	15.90	15.17	12.34	14.63	17.95	20.18
	POS [15]	N/A	3.67	7.25	4.00	3.86	4.51	6.12	12.36	14.43
SUPERVISED		UBFC-rPPG [3]	3.69	3.38	N/A	N/A	5.13	6.53	14.00	15.47
	TS-CAN [16]	PURE [9]	N/A	N/A	1.29	1.50	5.72	7.34	13.93	15.14
		SCAMPS [8]	4.66	5.83	3.62	3.53	5.55	6.91	19.05	21.77
		UBFC-rPPG [3]	9.36	17.84	N/A	N/A	5.51	7.47	10.23	12.46
	PhysNet [17]	PURE [9]	N/A	N/A	1.63	1.68	5.07	6.37	13.21	14.73
		SCAMPS [8]	20.08	31.27	4.39	4.39	7.28	9.98	21.05	24.69
		UBFC-rPPG [3]	5.54	5.32	N/A	N/A	6.62	8.21	17.49	19.26
	DeepPhys [18]	PURE [9]	N/A	N/A	1.21	1.42	8.42	10.18	16.92	18.54
		SCAMPS [8]	3.95	4.25	3.10	3.08	4.75	5.89	15.22	16.56
		UBFC-rPPG [3]	5.47	5.39	N/A	N/A	4.93	6.25	13.78	15.15
	EfficientPhys [19]	PURE [9]	N/A	N/A	2.07	2.10	5.31	6.61	14.03	15.31
		SCAMPS [8]	10.24	11.70	12.64	11.26	6.97	8.47	20.41	23.52

Supervised learning methods consistently outperform their unsupervised counterparts across most datasets, as evidenced by lower Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) values. Methods such as TS-CAN and Deep-Phys exhibit exceptional accuracy, particularly on the UBFC-rPPG dataset, underscoring the effectiveness of these advanced techniques in capturing the complex patterns inherent in rPPG signals.

For instance, TS-CAN achieves an MAE of 1.29 and an MAPE of 1.50 on the UBFC-rPPG dataset, while Deep-Phys records an MAE of 1.21 and an MAPE of 1.42 on the same dataset. These results demonstrate the precision and reliability of supervised methods in rPPG analysis.

Unsupervised methods, while advantageous for their independence from labelled training data, show higher variability and generally larger error margins. This is particularly evident in the performance of methods like GREEN and ICA, which exhibit significant errors across multiple datasets. For example, the GREEN method records an MAE of 19.81 and an MAPE of 18.78 on the UBFC-rPPG dataset, highlighting its relative inefficacy compared to supervised techniques.

The performance of the methods varies significantly across different datasets, emphasizing the importance of dataset characteristics in rPPG analysis.

- PURE Dataset: The POS and UBFC-rPPG supervised methods perform well, with POS achieving an MAE of 3.67 and an MAPE of 7.25.
- UBFC-rPPG Dataset: Supervised methods such as TS-CAN and Deep-Phys lead in performance, with notably low error rates.
- UBFC-Phys Dataset: Supervised methods again show superior performance, with PhysNet achieving an MAE of 5.07 and an MAPE of 6.37.
- MMPD Dataset: The performance trend continues, with supervised methods like PhysNet and UBFC-rPPG showing lower error rates.

A key finding from the data is the variability in the robustness and generalizability of different methods. Some methods, such as SCAMPS in various configurations, show inconsistent performance across datasets, suggesting potential overfitting or sensitivity to specific data characteristics. For example, SCAMPS paired with PhysNet records a high MAE of 20.08 and an MAPE of 31.27 on the PURE dataset, indicating significant performance degradation.

Conversely, methods like TS-CAN and Deep-Phys demonstrate robust performance across multiple datasets, indicating better generalizability. Their consistent low error rates across diverse datasets suggest that these methods are more resilient to variations in data, making them more reliable for real-world applications.

6. Summary

rPPG represents a promising approach for non-contact physiological monitoring with diverse applications in healthcare, human–computer interaction, and biometric identification. By leveraging advancements in signal processing and machine learning, rPPG methods continue to evolve, offering new opportunities for remote sensing and analysis of cardiovascular signals.

In conclusion, rPPG emerges as a promising methodology for non-contact physiological monitoring with wide-ranging applications in healthcare, human–computer interaction, and biometric identification. Leveraging advancements in signal processing and machine learning, rPPG methods continue to evolve, presenting novel avenues for remote sensing and analysis of cardiovascular signals.

Colour-based methods, exemplified by the GREEN, CHROM, and LGI techniques, offer robust solutions for cardiovascular signal extraction by capitalizing on changes in skin colour captured by video data. These methods demonstrate resilience to lighting variations, adaptability to diverse skin tones, and the convenience of contactless monitoring, opening new frontiers for non-invasive physiological monitoring across various domains.

Motion-based methods, such as the independent component analysis (ICA), Motion Magnification, and Spatial Filtering approaches, leverage motion information to extract cardiovascular signals from video data. With their robustness to motion artefacts, dynamic signal extraction capabilities, and contactless monitoring advantages, motion-based rPPG methods present promising avenues for non-invasive physiological monitoring in diverse applications.

Deep learning approaches, including 3D Convolutional Networks (3D CNNs), offer powerful frameworks for the automatic extraction of cardiovascular signals from video data. With their end-to-end learning capabilities, robustness to variability, and scalability advantages, deep learning models provide new opportunities for non-contact physiological monitoring across various domains and applications.

Multispectral methods, near-infrared imaging, hemoglobin spectroscopy, Dual-Wavelength Imaging, and depth-based methods, each offer unique advantages in capturing and analyzing cardiovascular signals. These methods leverage diverse sensing modalities, such as spectral analysis, near-infrared light, and depth sensing technologies, to enhance the robustness and accuracy of signal extraction, paving the way for non-invasive physiological monitoring in various applications.

Hybrid Models represent a promising direction in rPPG, integrating multiple approaches to enhance the accuracy, robustness, and versatility of cardiovascular signal extraction from video data. By combining complementary strengths of different methods, Hybrid Models offer new opportunities for non-invasive physiological monitoring in diverse domains and applications, driving innovation in remote sensing and analysis of cardiovascular signals.

7. Conclusions

The results emphasize the need for additional research to enhance the effectiveness and resilience of unsupervised methods. Investigating hybrid methodologies that integrate the advantages of both supervised and unsupervised paradigms may present promising opportunities for progress. Moreover, the inclusion of a wide range of training datasets is essential for the development of resilient models. It is crucial to have diverse and inclusive datasets in order to make progress in rPPG technology and attain dependable, practical uses.

To summarize, supervised learning methods currently outperform other methods in analyzing rPPG. However, further innovation and thorough evaluation using various datasets are crucial for advancing the field and creating rPPG methods that are accurate, reliable, and applicable to different scenarios.

Author Contributions: Conceptualization, L.S. and V.R.; methodology, V.R.; software, L.S.; formal analysis, G.Š.; investigation and writing—original draft preparation, L.S.; writing—review and editing, A.K.-G.; visualization, G.Š. and A.K.-G.; supervision, V.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-funded by the European Union under Horizon Europe programme grant agreement No. 101059903, by the European Union funds for the period 2021–2027, and by the state budget of the Republic of Lithuania financial agreement Nr. 10-042-P-0001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hertzman, A.B. Photoelectric Plethysmography of the Nasal Septum in Man. *Proc. Soc. Exp. Biol. Med.* **1937**, *37*, 290–292. [CrossRef]
2. Millikan, G.A. The Oximeter, an Instrument for Measuring Continuously the Oxygen Saturation of Arterial Blood in Man. *Rev. Sci. Instrum.* **1942**, *13*, 434–444. [CrossRef]
3. Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote Plethysmographic Imaging Using Ambient Light. *Opt. Express* **2008**, *16*, 21434–21445. [CrossRef] [PubMed]
4. de Haan, G.; Jeanne, V. Robust Pulse Rate from Chrominance-Based RPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [CrossRef] [PubMed]
5. Pilz, C.S.; Zaunseder, S.; Krajewski, J.; Blazek, V. Local Group Invariance for Heart Rate Estimation from Face Videos in the Wild. *IEEE Xplore.* **2018**, 1367–1375. [CrossRef]
6. de Haan, G.; van Leest, A. Improved Motion Robustness of Remote-PPG by Using the Blood Volume Pulse Signature. *Physiol. Meas.* **2014**, *35*, 1913–1926. [CrossRef] [PubMed]
7. Wang, W.; Den Brinker, A.C.; Stuijk, S.; De Haan, G. Algorithmic Principles of Remote PPG. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1479–1491. [CrossRef] [PubMed]
8. Casado, C.Á.; López, M.B. Face2PPG: An Unsupervised Pipeline for Blood Volume Pulse Extraction from Faces. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 5530–5541. [CrossRef] [PubMed]
9. Poh, M.-Z.; McDuff, D.J.; Picard, R.W. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Trans. Biomed. Eng.* **2010**, *58*, 7–11. [CrossRef]
10. Wu, H.-Y.; Rubinstein, M.; Shih, E.; Guttag, J.; Durand, F.; Freeman, W. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Trans. Graph.* **2012**, *31*, 65. [CrossRef]
11. 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.* **2022**, *14*, 1630–1641. [CrossRef]
12. Przybyło, J. A Deep Learning Approach for Remote Heart Rate Estimation. *Biomed. Signal Process. Control* **2022**, *74*, 103457. [CrossRef]
13. Kado, S.; Monno, Y.; Yoshizaki, K.; Tanaka, M.; Okutomi, M. Spatial-Spectraloral Fusion for Remote Heart Rate Estimation. *IEEE Sens. J.* **2020**, *20*, 11688–11697. [CrossRef]
14. Ouzar, Y.; Djeldjli, D.; Bousefsaf, F.; Maaoui, C. X-IPPGNet: A Novel One Stage Deep Learning Architecture Based on Depthwise Separable Convolutions for Video-Based Pulse Rate Estimation. *Comput. Biol. Med.* **2023**, *154*, 106592. [CrossRef] [PubMed]

15. Firmansyah, R.A.; Prabowo, Y.A.; Suheta, T.; Muharom, S. Implementation of 1D Convolutional Neural Network for Improvement Remote Photoplethysmography Measurement. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *29*, 1326–1335. [CrossRef]
16. McDuff, D.; Wander, M.; Liu, X.; Hill, B.L.; Hernandez, J.; Lester, J.; Baltrusaitis, T. SCAMPS: Synthetics for Camera Measurement of Physiological Signals. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 3744–3757.
17. Bobbia, S.; Macwan, R.; Benezeth, Y.; Mansouri, A.; Dubois, J. Unsupervised Skin Tissue Segmentation for Remote Photoplethysmography. *Pattern Recognit. Lett.* **2017**, *124*, 82–90. [CrossRef]
18. Stricker, R.; Steffen Müller, S.; Gross, H.-M. Non-Contact Video-Based Pulse Rate Measurement on a Mobile Service Robot. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; pp. 1056–1062.
19. Li, X.; Wang, T.; Zhao, G.; Zhang, X.; Kang, X.; Yin, L. ReactioNet: Learning High-Order Facial Behavior from Universal Stimulus-Reaction by Dyadic Relation Reasoning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023.
20. Meziati, R.; Benezeth, Y.; De Oliveira, P.; Chappé, J.; Yang, F. UBFC-Phys 2021. Available online: https://ieee-dataport.org/open-access/ubfc-phys-2 (accessed on 20 December 2022).
21. Tang, J.; Chen, K.; Wang, Y.; Shi, Y.; Patel, S.; McDuff, D.; Liu, X. MMPD: Multi-Domain Mobile Video Physiology Dataset. In Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 24–27 July 2023; pp. 1–5.
22. Joshi, J.; Cho, Y. IBVP Dataset: RGB-Thermal RPPG Dataset with High Resolution Signal Quality Labels. *Electronics* **2024**, *13*, 1334. [CrossRef]
23. Zhang, Z.; Girard, J.M.; Wu, Y.; Zhang, X.; Liu, P.; Ciftci, U.; Canavan, S.; Reale, M.; Horowitz, A.; Yang, H.; et al. Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3438–3446.
24. Wang, D.; Chahl, J. Simulating Cardiac Signals on 3D Human Models for Photoplethysmography Development. *Front. Robot. AI* **2023**, *10*, 1266535. [CrossRef] [PubMed]
25. Liu, X.; Narayanswamy, G.; Paruchuri, A.; Zhang, X.; Tang, J.; Zhang, Y.; Sengupta, S.; Patel, S.; Wang, Y.; McDuff, D. RPPG-Toolbox: Deep Remote PPG Toolbox. *Adv. Neural Inf. Process. Syst.* **2022**, *36*, 68485–68510.