

Review

RGB-Only rPPG for Contactless SpO₂ Estimation: A Critical Review of Optical Constraints, Calibration Dependence, and Clinical Validation

Sajana Islam Moon and Nam Kyu Kwon * 

Department of Electronic Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea; sajanaislam@yu.ac.kr

* Correspondence: namkyu@yu.ac.kr; Tel.: +82-53-810-3095

Abstract

This review focuses exclusively on RGB camera-based remote photoplethysmography (rPPG) for contactless estimation of peripheral oxygen saturation (SpO₂). Unlike heart-rate estimation, which mainly depends on recovering pulse periodicity, SpO₂ estimation requires oxygenation-sensitive optical contrast, stable channel-wise AC/DC information, and calibration and validation against a physiological reference. The reviewed literature is organized around the main stages of an RGB-only SpO₂ pipeline. Existing methods are further discussed as ratio-based or Beer–Lambert-inspired approaches, classical regression models, deep learning and spatiotemporal methods, and hybrid or physics-informed designs. This critical review shows that RGB-based systems can estimate oxygenation-related signals in controlled settings. However, current evidence is limited by device-dependent camera responses, narrow oxygen saturation ranges, weak validation, limited demographic reporting, and insufficient hypoxemia testing. The main conclusion of this review is that RGB-only SpO₂ estimation is useful under controlled conditions, but it should not be treated as an absolute, clinically interpretable SpO₂ measurement method without stronger validation. Future work should focus on wider SpO₂ distributions, transparent calibration procedures, cross-device testing protocols, subgroup reporting, and a clear distinction between oxygenation trend monitoring and clinically interpretable absolute SpO₂ measurement.

Keywords: RGB camera; remote photoplethysmography; contactless SpO₂ estimation; oxygen saturation; pulse oximetry; calibration; clinical validation



Academic Editors: John A. Kalomiros, Cheng-Ying Chung, Yu-Feng Chung and Chang-Wei Hsieh

Received: 29 June 2026

Revised: 31 July 2026

Accepted: 4 August 2026

Published: 6 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

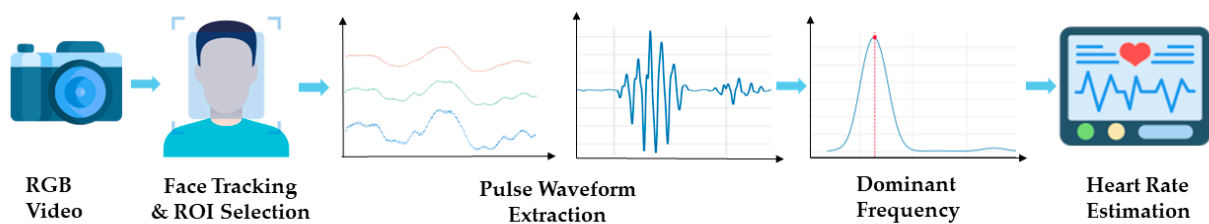
Pulse oximetry is a standard method for monitoring peripheral oxygen saturation (SpO₂) because it provides a clinically useful physiological estimate from optical measurements [1–4]. In a traditional contact pulse oximeter, selected red and near-infrared light is transmitted or reflected through tissue, where oxygenated and deoxygenated hemoglobin show different wavelength-dependent absorption behavior [1,2,5,6]. The pulsatile arterial component is separated from the non-pulsatile background, and the resulting ratio-of-ratios is mapped to SpO₂ through empirical calibration [1,5,7]. From this workflow, it is clear that SpO₂ measurement is not only a pulse-detection task; it also depends on wavelength selection, optical stability, signal separation, and calibration against a physiological reference.

Remote photoplethysmography (rPPG) changes the measurement condition because physiological information is inferred from small skin-color variations in video rather than

from a contact optical probe [8,9]. This makes RGB camera-based monitoring attractive for contactless applications, especially where wearable or contact sensors are inconvenient [10–13]. Despite this potential, RGB-only rPPG does not measure the same optical signal as pulse oximetry. A consumer RGB camera records broad, device-dependent visible color channels rather than controlled red and near-infrared wavelength bands [14–16]. These channels are affected by illumination, camera response, skin optics, motion, and acquisition-related variation [15–18].

This difference is important insofar as SpO₂ estimation is more demanding than heart-rate estimation. Heart-rate rPPG mainly depends on recovering the timing of the pulsatile waveform, and several rPPG methods were originally developed for pulse or heart-rate recovery from video [8,9,19]. If the pulsatile component is sufficiently visible, the dominant frequency can often provide a useful heart-rate estimate. SpO₂ estimation, however, requires more than pulse timing. It needs oxygenation-dependent optical contrast between channels, together with stable amplitude and baseline information [5,7]. A video may contain a strong rPPG pulse but still be unreliable for absolute SpO₂ estimation if the channel-wise AC/DC relationship is unstable [7,14]. This distinction is illustrated in Figure 1, which shows that heart-rate rPPG mainly depends on pulse periodicity, whereas SpO₂-oriented rPPG requires channel-dependent AC/DC information and calibration.

(a) HR-oriented rPPG pipeline



(b) SpO₂-oriented rPPG pipeline

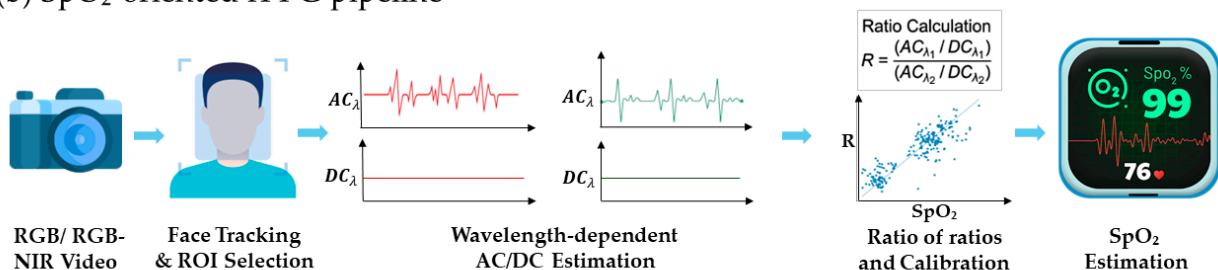


Figure 1. Conceptual difference between (a) heart-rate remote photoplethysmography (HR-rPPG) and (b) SpO₂-oriented rPPG.

RGB-only SpO₂ studies have used several methodological strategies, including ratio-of-ratios or calibrated RGB approaches [15,16,20], classical feature-based or regression methods [21–23], and deep learning or spatiotemporal representation-based methods [24–27]. These studies show that RGB-only methods can produce useful oxygenation-related estimates under controlled or constrained settings. The dominant limitation is that RGB camera channels have limited spectral specificity compared with conventional red/NIR pulse oximetry [7,14,15]. A model may reduce noise or learn an empirical mapping from RGB-derived features, but it cannot fully replace missing wavelength information if the camera did not capture it in a stable and physiologically meaningful way. For this reason, reported error values must be interpreted together with experimental design, acquisition conditions, validation split, skin-tone diversity, and clinical agreement analysis [28–32].

Our review focuses on remote SpO₂ estimation using RGB-camera-derived rPPG signals. Other sensing modalities, such as RGB-NIR, multispectral, hyperspectral, and controlled spectral-imaging systems, are discussed only when they help explain the limitations of RGB-only approaches [33–37]. We evaluated existing RGB-only methods in terms of spectral ambiguity, signal processing, calibration dependence, dataset composition, validation protocol, and clinical claim strength. Unlike broader surveys, it specifically examines whether RGB cameras can provide reliable absolute SpO₂ estimation despite limited spectral specificity and weak low-saturation validation.

Several recent reviews have addressed topics closely related to camera-based SpO₂ estimation, but their scope and evaluation focus differ from the present review. Pirzada et al. [10] reviewed rPPG methods for both heart-rate and blood-oxygenation measurement, with emphasis on the processing pipeline from ROI selection and signal extraction to preprocessing and vital-sign estimation. Saikevičius et al. [11] examined non-contact vision-based monitoring across multiple vital signs, while Hassanpour and Yang [12] considered a broader multimodal and multitask framework that includes vision, radar, thermal imaging, and other sensing modalities. Sakib et al. [13] similarly surveyed rPPG as a general platform for several physiological parameters, including heart-rate, respiratory rate, and SpO₂. More recently, Wuyart et al. [38] provided a broad survey of video-based blood-oxygen-saturation assessment. These reviews establish the broader development of contactless physiological monitoring, but they do not make the measurement validity of RGB-only SpO₂ the central question.

The contribution of the present review is therefore not simply another classification of rPPG algorithms; it isolates RGB-only SpO₂ estimation as a distinct measurement problem and examines whether broad, device-dependent visible camera channels preserve sufficient oxygenation-sensitive information for calibrated and clinically interpretable SpO₂ estimation. The reviewed methods are evaluated jointly according to optical constraints, calibration dependence, signal representation, validation design, low-saturation coverage, cross-condition generalization, and the strength of the clinical claim supported by the available evidence. Table 1 summarizes this distinction from the broader review literature.

Table 1. Position of this review compared with existing surveys.

Review Type	Main Focus	Limitation for RGB-Only SpO ₂	Difference of This Review
General rPPG reviews	Video-based pulse and vital-sign monitoring	SpO ₂ is not treated as a separate measurement problem	Focuses on RGB-only SpO ₂
Contactless vital sign reviews	Multi-parameter non contact monitoring	Limited focus on spectral ambiguity and low-SpO ₂ validation	Evaluates absolute SpO ₂ claim strength
Video-based blood oxygenation reviews	Imaging-based oxygenation assessment	Often includes RGB-NIR, multispectral or hyperspectral systems	Separates RGB-only evidence from non-RGB comparison
This review	RGB-only camera-based SpO ₂ estimation	Not applicable	Assesses optical limits, calibration, validation and clinical claim level

The central conclusion developed in this review is that RGB-only rPPG can support oxygenation-related estimation under constrained conditions, but clinically interpretable absolute SpO₂ claims require stronger evidence, especially calibration, low-saturation validation, agreement analysis, and generalization across subjects, devices, and skin-tone groups. To support this argument, the review proposes a validation-oriented framework that separates pulse recovery from oxygenation measurement, organizes methods by sensing constraint, signal representation, estimation principle, calibration requirement, and

validation strength, and links each method category to the level of claim supported by the available evidence.

2. Review Methodology

This review is designed as a critical narrative review with structured study extraction. The focus was RGB-only camera-based SpO₂ estimation. The aim was not to perform a statistical meta-analysis. We examined whether existing RGB-only methods provide sufficient optical, signal-processing, calibration and validation evidence to support their claimed SpO₂ estimation accuracy. Non-RGB camera-based studies were used only as comparison cases when they helped us to explain the limitations of RGB-only sensing.

2.1. Literature Search

In this review, RGB-only SpO₂ estimation includes methods using raw RGB video, RGB traces, RGB-derived rPPG, patch-level signals, or RGB-based spatiotemporal representations, provided that the final SpO₂ estimate does not require an additional NIR, multispectral, hyperspectral, thermal, or non-camera sensing channel.

Relevant studies were searched in PubMed, IEEE Xplore, ScienceDirect and Google Scholar. Searches were last updated on 11 June 2026. The search terms included combinations of “remote photoplethysmography SpO₂,” “rPPG oxygen saturation,” “RGB camera SpO₂,” “contactless SpO₂,” “non-contact pulse oximetry,” and “imaging photoplethysmography oxygen saturation”. Backward and forward citation checking was also performed for key papers that were repeatedly cited in contactless SpO₂ research.

Candidate studies were assessed according to the sensing modality, target variable, availability of reference SpO₂, and presence of experimental validation. A study was classified as core RGB-only evidence when SpO₂ was explicitly estimated from RGB video, RGB skin traces, RGB-derived rPPG, or an RGB-based spatiotemporal representation without requiring NIR or additional spectral channels for the SpO₂ estimate. Studies using RGB-NIR, multispectral, hyperspectral, or other additional sensing modalities were not counted as core RGB-only evidence, although relevant studies were retained as comparison evidence when they clarified spectral, calibration, or validation limitations. Studies that only focused on heart-rate, respiratory rate, blood pressure, general rPPG extraction, or other physiological variables without SpO₂ estimation were excluded from the core evidence set.

The core RGB-only evidence set comprised 13 experimental studies directly addressing RGB video-based SpO₂ estimation. A further 11 non-RGB studies using NIR, RGB-NIR, dual-wavelength, multispectral, hyperspectral, reflectance-based, or related spectral approaches were retained for comparison and methodological context. The core RGB-only evidence base consisted of studies [14–16,20–28,39]. Existing reviews, general rPPG studies, pulse oximetry fundamentals, bias studies, and regulatory documents were treated as the supporting literature and were not included in these counts.

The selection was not restricted to a single algorithm family. Ratio-of-ratios methods, Beer–Lambert-inspired approaches, classical regression, machine learning, deep learning, and hybrid methods were considered when they explicitly estimated SpO₂ or oxygen-saturation-related information from RGB video, RGB skin traces, RGB-derived rPPG signals, or RGB-based spatiotemporal representations. In the representative-study comparison tables, studies using RGB-NIR, multispectral, hyperspectral, thermal, contact PPG or wearable sensors were not treated as core RGB-only evidence. These studies were used only as comparison studies to clarify the sensing limitations of RGB-only SpO₂ estimation.

2.2. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for the core RGB-only evidence base are summarized in Table 2.

Table 2. Inclusion and exclusion criteria used for literature selection.

Category	Inclusion Criteria	Exclusion Criteria
Target variable	SpO ₂ , hypoxemia, or oxygenation-related estimation	HR, RR, BP, or general rPPG only
Sensing modality	RGB video, RGB traces, RGB-rPPG, or RGB spatiotemporal maps	Contact PPG, wearable, radar, thermal, or non-camera sensing
RGB-only scope	Main input limited to RGB channels	RGB-NIR, multispectral, hyperspectral, or spectral imaging as core evidence
Methodological relevance	Ratio-based, Beer-Lambert-inspired, regression, ML, DL, transformer, or hybrid SpO ₂ methods	Vision/rPPG methods without SpO ₂ estimation
Reference measurement	Reference SpO ₂ from pulse oximetry or physiological measurement	No reference or labeled SpO ₂ available for evaluating the estimation method
Experimental evidence	Experimental validation, dataset testing, benchmarking, or clinical measurement analysis	Conceptual, review, or methodological papers without experimental SpO ₂ evaluation
Validation information	Reported subjects, setup, SpO ₂ range, dataset, or metrics	Validation design cannot be interpreted
Comparison evidence	Non-RGB studies used only for explaining RGB limitations	Non-RGB studies treated as RGB-only evidence

2.3. Data Extraction

For each included RGB-only study, the following information was extracted: sensing setup, camera type when reported, anatomical region or region of interest, RGB signal representation, preprocessing method, AC/DC or feature extraction strategy, estimation model, calibration strategy, reference SpO₂ source, subject number, SpO₂ range, validation split, evaluation metrics, and stated limitations. When a study did not report a required item, it was marked as “not reported”.

The extracted information was used to determine which method and what level of claim the validation could support. In this review, we emphasize subject-independent testing, low-saturation coverage, cross-device or cross-dataset validation and skin-tone variation. Information was extracted at the study level rather than pooled statistically because the acquisition and validation protocols were not comparable across studies.

2.4. Structured Evidence Appraisal

To reduce reliance on purely narrative judgment, the core RGB-only studies were assessed using a structured evidence-appraisal framework. A single numerical quality score was not used because the included studies differed substantially in sensing design, estimation principle, calibration strategy, and validation objective. Instead, each study was evaluated across predefined domains that directly affect the interpretation of RGB-only SpO₂ performance: participant and dataset characteristics, SpO₂ range and hypoxemia coverage, reference SpO₂ measurement, subject/session independence of the validation split, calibration independence, cross-device or cross-dataset evaluation, demographic and skin-tone reporting, and the use of clinically interpretable error or agreement metrics.

For each domain, the available evidence was classified as adequate, partial, unclear/not reported, or not demonstrated. These categories were used to support the validation strength interpretation presented later in this review rather than to generate a pooled numerical quality score. The appraisal approach was informed by general principles of prediction-model risk-of-bias assessment, particularly the importance of participant selection, predictor measurement, outcome measurement, and analysis design [32]. Because the core evidence also includes optical and ratio-based methods that are not conventional prediction models, the appraisal framework was adapted to the measurement-specific limitations of RGB-only SpO₂ estimation rather than applying a prediction-model tool unchanged.

The operational criteria used for this appraisal are provided in Supplementary Table S1. Each domain was classified as adequate, partial, unclear/not reported, or not demonstrated according to predefined criteria. These domain ratings were not summed into a numerical quality score. Instead, the validation claim level was treated as an evidence ceiling: stronger claims required progressively stronger evidence for participant independence, calibration independence, low-SpO₂ coverage, cross-domain generalization, demographic reporting, and clinical agreement. Consequently, strong performance in one domain, such as a low MAE, was not considered sufficient to compensate for the absence of an essential validation requirement. Supplementary Table S1 summarizes the resulting appraisal of the 13 core RGB-only studies.

2.5. Review Framework

The studies that we included were analyzed using four linked questions. First, what optical information can the RGB-only sensing setup realistically capture? Second, how does the signal-processing pipeline preserve or distort channel-wise amplitude and baseline information? Third, what calibration strategy is used to map RGB-derived features to SpO₂? Finally, what level of clinical claim is supported by the validation design?

Based on these questions, the selected studies were organized by sensing limitations, validation quality, and estimation methodology. Evidence strength was interpreted across five levels: feasibility-level evidence, constrained laboratory estimation, subject-independent estimation, cross-condition generalization, and clinically relevant hypoxemia validation. This structure was adopted because average error values alone do not show whether a method is robust across subjects, devices, lighting conditions, skin tones, or low-SpO₂ ranges.

2.6. Synthesis Approach

The findings were synthesized qualitatively. A statistical meta-analysis was not performed because the studies differed in camera hardware, lighting condition, ROI selection, preprocessing, reference SpO₂ source, saturation range, validation split, and evaluation metrics. These differences make direct quantitative pooling unreliable. Instead, the review focuses on recurring technical limitations and validation gaps, including spectral ambiguity, camera-pipeline dependence, calibration sensitivity, narrow saturation range bias, subject or session leakage, weak low-SpO₂ testing, and limited clinical agreement analysis. The available RGB-only SpO₂ studies differ substantially in sensing setup, calibration strategy, validation protocol and reported metrics; accordingly, the synthesis should be interpreted as a structured critical review rather than as a quantitative systematic meta-analysis. Because this study was not designed as a formal systematic review, the 13-study core evidence set should not be interpreted as an exhaustive inventory of every publication mentioning RGB-derived oxygen saturation.

3. Optical and Physiological Basis of SpO₂ Estimation

A useful way to understand RGB-only SpO₂ estimation is to separate pulse detection from oxygenation measurement. Heart-rate estimation mainly asks whether the periodic pulse rhythm can be recovered from the video signal. SpO₂ estimation asks a different question: whether the system can measure an oxygenation-dependent optical contrast. Although heart-rate rPPG is mainly a temporal problem, SpO₂ estimation is both temporal and spectral [1,7].

Conventional pulse oximetry estimates arterial oxygen saturation from the different optical absorption properties of oxygenated hemoglobin (HbO₂) and deoxygenated hemoglobin (Hb). Red and near-infrared wavelengths are commonly used because the relative absorption of HbO₂ and Hb changes across these spectral regions [40]. The device separates the pulsatile arterial component from the non-pulsatile background, calculates normalized pulsatile changes at two wavelengths, and maps the resulting ratio to SpO₂ using empirical calibration [41]. The wavelength pair and calibration curve are therefore central parts of the measurement, not minor implementation details.

The measured optical signal at a wavelength λ can be written in simplified form as

$$S_{\lambda}(t) = DC_{\lambda} + AC_{\lambda}(t) \quad (1)$$

where $S_{\lambda}(t)$ is the measured optical signal at time t , $DC_{\lambda}(t)$ is the slowly varying or baseline component, and $AC_{\lambda}(t)$ is the pulsatile component associated mainly with arterial blood volume changes. The DC_{λ} term includes contributions from tissue, non-pulsatile blood, illumination, optical path, and device response. Pulse oximetry uses the normalized pulsatile component rather than the raw signal amplitude:

$$R = \frac{\frac{AC_{\lambda_1}}{DC_{\lambda_1}}}{\frac{AC_{\lambda_2}}{DC_{\lambda_2}}} \quad (2)$$

For conventional pulse oximetry, the two wavelengths are usually represented by red and near-infrared channels:

$$R = \frac{\frac{AC_{red}}{DC_{red}}}{\frac{AC_{IR}}{DC_{IR}}} \quad (3)$$

SpO₂ is then estimated using an empirical calibration function:

$$SpO_2 = f(R) \quad (4)$$

For explanation, this relationship is sometimes written in a simplified linear form:

$$SpO_2 = A - BR \quad (5)$$

where A and B are calibration coefficients. This linear form should not be interpreted as a universal physiological law. The calibration relationship depends on wavelength selection, sensor design, optical geometry, tissue scattering, device response, and the population used for calibration [5,7,18].

Equations (1)–(5) are simplified representations intended to explain the pulse oximetry principle and should not be interpreted as a direct application of the Beer–Lambert law to RGB camera measurements. Biological tissue is heterogeneous and strongly scattering, so the effective optical path length depends on wavelength, tissue composition, measurement geometry, and scattering properties [5,18]. The measured reflectance is also affected by melanin, surface reflection, and other tissue constituents. In addition, an RGB camera does

not measure light at a single wavelength. Each color channel integrates a broad range of wavelengths according to the illumination spectrum and the camera spectral sensitivity, and the RGB sensitivity functions can overlap substantially and vary between devices [42]. Therefore, Beer–Lambert-inspired RGB methods should be regarded as approximate or empirically calibrated models rather than direct spectrophotometric measurements of hemoglobin absorption.

This point is especially important for RGB-only contactless SpO₂ estimation. In a conventional pulse oximeter, the red and near-infrared channels are controlled optical measurements. In an RGB camera, the red, green, and blue channels are broad camera responses shaped by illumination, sensor spectral sensitivity, lens properties, camera processing, and skin optics [14–16]. A ratio can still be calculated from RGB traces, but its physiological meaning is less stable than a conventional red/NIR ratio, as the effective RGB channel responses depend on illumination, sensor spectral sensitivity, camera processing, and skin optics.

Methodological Necessity of System Calibration

Calibration is necessary given the influence of tissue scattering, optical path length, illumination spectrum, sensor response, skin pigmentation and subject-specific physiological differences on the measured optical ratio. For RGB-only cameras, this problem is compounded by the fact that the camera channels are not controlled red and near-infrared wavelength bands; they are broad visible channels whose effective response can change with lighting, camera settings and skin properties [15,16,18].

Therefore, calibration is the step that links RGB-derived features or ratios to reference SpO₂ values. Without this step, an RGB-derived ratio should be interpreted only as an oxygenation-related optical feature; it cannot be treated as an absolute SpO₂ measurement. Even after calibration, the claim should remain limited to the conditions that have been tested in the study. A calibration obtained under one camera, lighting condition, subject group or SpO₂ range may not remain valid under different acquisition conditions unless this generalization is directly evaluated.

4. rPPG Signal Formation and Processing Pipeline

An SpO₂ estimation pipeline differs from a heart-rate rPPG pipeline in that it must preserve channel-dependent amplitude and baseline relationships. Many preprocessing operations that improve heart-rate estimation may alter the amplitude and baseline relationships that are needed for AC/DC ratio analysis. Consequently, a heart-rate rPPG method that performs well for pulse detection does not automatically provide reliable SpO₂ estimation.

4.1. ROI Detection and Tracking

ROI selection determines which skin region is used to extract the RGB traces. ROI is mostly selected from facial skin. In some studies, hand regions or smaller patch-based areas are also used [16,21,24,25]. For heart-rate estimation, ROI is often judged mainly by whether it contains a clear pulsatile rhythm [43]. For SpO₂ estimation, this is not sufficient. The selected region also affects the relative AC and DC components of each RGB channel [14].

This makes ROI stability important. Small changes in ROI location can change the mixture of skin tissue, blood perfusion, facial hair and local pigmentation. A moving or inconsistent ROI may still preserve the pulse frequency, but it can distort the channel-wise amplitude and baseline relationship. This information is crucial for estimating oxygen saturation. RGB-only SpO₂ studies should report the anatomical ROI, tracking method, skin segmentation strategy and any quality-control step used to reject unstable windows.

4.2. RGB Trace Extraction

After ROI selection, most RGB-based rPPG pipelines convert the selected skin pixels into RGB signals. This is usually done by averaging pixel values within the ROI for each video frame. Some recent methods use patch-level traces or spatiotemporal maps [24,26]. Spatial averaging can improve the visibility of the pulse signal by reducing random pixel-level noise. It also mixes regions with different physiological, optical and motion characteristics [44,45]. This mixing is sometimes acceptable for heart-rate estimation as long as the periodic pulse component is still clear.

For SpO₂ estimation, oxygen saturation depends on the relative amplitude and baseline of the RGB channels, making this a particularly sensitive issue. If the extracted signals combine skin areas with different optical properties, it can distort the AC and DC components which are important for oxygen saturation measurement.

Using patch-based or spatiotemporal representations can preserve more local details. However, they may also increase the chance that a model learns irrelevant patterns, such as subject appearance, lighting configuration, or dataset-specific visual cues, instead of oxygenation-related information [17,25,26,32]. RGB trace extraction should be reported with enough detail. This includes the ROI averaging method, patch strategy, color normalization, and the rejection of unstable or low-quality regions.

4.3. Temporal Preprocessing

This step in the rPPG pipeline includes detrending, normalization, bandpass filtering, resampling, and windowing. These steps are useful for extracting pulsatile signals from raw RGB signals. They reduce unwanted components such as motion artifacts, camera noise, sudden lighting changes and baseline drift [43,46–49]. For SpO₂, strong normalization may reduce lighting variation while also removing baseline information needed for AC/DC estimation. Similarly, filter bandwidth and window length can change the estimated pulse amplitude even when the physiological saturation is unchanged. This sensitivity is visible in both ratio-of-ratios methods and calibrated RGB-camera approaches. Therefore, these steps in temporal preprocessing must be examined carefully.

4.4. Color-Space Projection Methods

Color-space projection and source-separation methods, including CHROM, POS, PCA, ICA, and PBV-related approaches, have been used to improve pulse waveform extraction under motion and illumination changes [9,19,49]. These methods are valuable for heart-rate estimation and may improve rPPG signal quality before SpO₂ estimation [47]. Their direct use for absolute SpO₂ estimation requires caution. SpO₂ estimation not only depends on pulse visibility, but also on wavelength- or channel-dependent amplitude and baseline relationships [7,14].

Projection methods often suppress intensity variations or combine RGB channels to isolate pulsatility. This can improve rhythm detection, but it may also remove or reshape the channel-specific AC/DC information used for SpO₂ estimation [14–16]. So, when color-space projection methods are used before SpO₂ estimation, the study should explain how oxygenation-sensitive information is preserved. If the original RGB channel relationships are altered, the method should justify whether the SpO₂ information is recalibrated, recovered through learning, or estimated from an alternative representation. Otherwise, improved pulse extraction should not be interpreted as evidence of reliable absolute SpO₂ estimation.

4.5. AC/DC Estimation

AC/DC estimation is one of the most sensitive stages in an RGB-only SpO₂ pipeline. In ratio-based methods, the AC component represents the pulsatile part of the signal, while the DC component represents the local baseline intensity. The SpO₂-related feature is usually derived from the normalized pulsatile change in each channel rather than from the raw RGB amplitude alone [7].

Different studies may estimate AC in different ways, such as peak-to-peak amplitude, spectral amplitude, RMS energy, envelope-based amplitude, or learned temporal features. The DC component is commonly estimated from the local mean or slowly varying baseline of the same channel. These choices are not interchangeable, because changing the AC definition, window length, filtering method or baseline estimation can change the final ratio even when the underlying physiological SpO₂ is unchanged [16,26].

This issue is more serious in RGB-only video than in controlled pulse oximetry. Motion, illumination changes, automatic exposure, white balance, compression, and weak perfusion can all affect the estimated AC and DC values [14,15]. A window may contain a clear pulse rhythm but still produce an unstable channel ratio. This is why RGB-only SpO₂ studies should clearly report how AC and DC are defined, which channels are used, how windows are selected, and whether signal-quality checks are applied before estimating SpO₂. Without this information, the reported SpO₂ value cannot be reliably interpreted or reproduced.

5. RGB-Only rPPG Is Not Equivalent to Pulse Oximetry

The central limitation of RGB-only SpO₂ estimation is not that the camera signal is always unusable; it is the lack of controlled spectral specificity. This creates failure modes that can look like ordinary noise, but the deeper problem is structural: the RGB signal is not a direct measurement of selected hemoglobin-sensitive wavelengths. The main uncertainty pathway is summarized in Figure 2, where illumination, skin optics, camera response, camera processing, RGB trace formation, and AC/DC extraction all contribute to uncertainty before the final SpO₂ prediction.

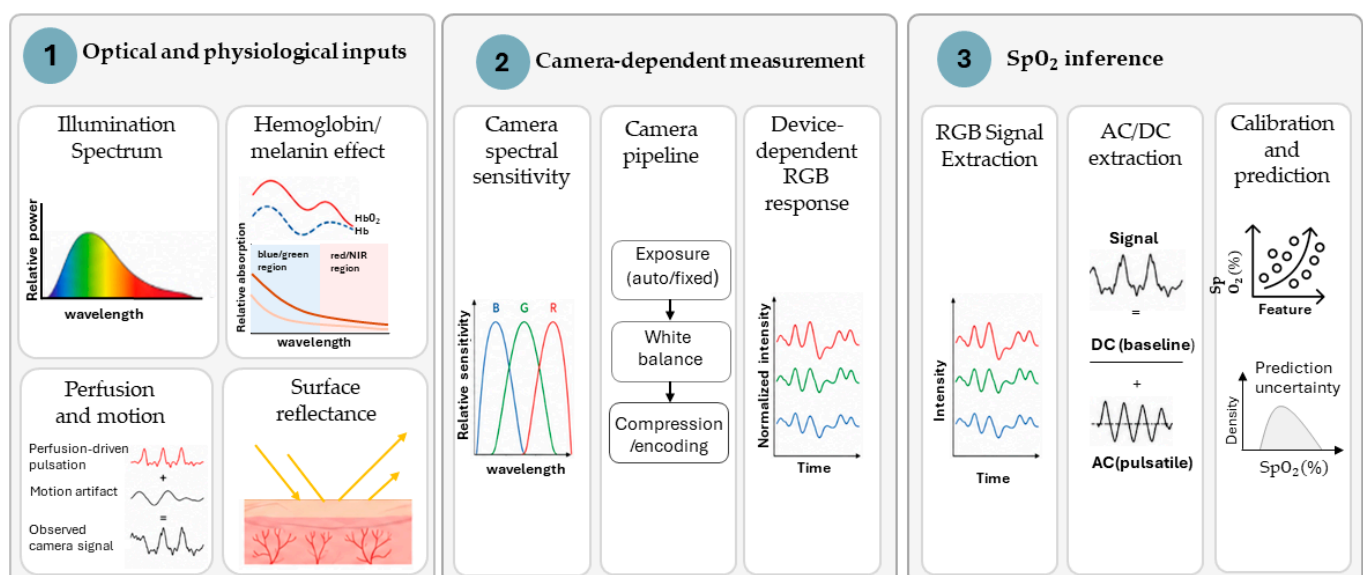


Figure 2. Sources of uncertainty in RGB-only SpO₂ estimation. Optical input, camera-dependent measurement, and signal-processing effects can distort the RGB traces and AC/DC ratio used for SpO₂ prediction.

5.1. Spectral Ambiguity

Unlike a narrowband optical measurement, an RGB camera channel represents a wavelength-integrated response determined jointly by the illumination spectrum, tissue reflectance, and sensor spectral sensitivity [42]. Because the RGB sensitivity functions are broad and partially overlapping, the red, green, and blue outputs should be interpreted as mixed spectral measurements rather than as measurements at three discrete wavelengths. A simplified expression for the measured camera channel k is

$$C_k(t) = \int E(\lambda) R_{skin}(\lambda, t) Q_k(\lambda) d\lambda \quad (6)$$

Here, $C_k(t)$ is the measured camera channel, $E(\lambda)$ is the illumination spectrum, $R_{skin}(\lambda, t)$ is the time-varying skin reflectance, and $Q_k(\lambda)$ is the camera spectral sensitivity. This equation shows why RGB values are not direct wavelength measurements. Even if the red channel contains oxygenation-related information, it also contains the effects of the illumination spectrum, sensor response, skin pigmentation, surface reflection, and camera processing [15,16]. Because the measured RGB signal is jointly determined by lighting conditions, skin reflectance, and sensor-dependent wavelength response, the effect of oxygenation cannot be isolated from RGB values alone without assumptions, calibration, or controlled acquisition [7,14]. This is a central reason why RGB-only SpO₂ estimation is less identifiable than red/NIR pulse oximetry. Modified camera designs, such as notch RGB-camera systems, have been explored to improve spectral separation for contactless SpO₂ estimation, including clinical testing in neonatal intensive care settings [50]. So, these approaches should be interpreted as comparison evidence rather than ordinary RGB-only evidence.

Conventional pulse oximetry typically uses two deliberately selected wavelength bands with red illumination near 660 nm and near-infrared illumination near 940 nm [1,41]. In contrast, the R, G, and B channels of a conventional color camera are broad, overlapping spectral sensitivity functions rather than measurements at single wavelengths. For example, the RGB camera characterized by van Gastel and Verkruyse [14] showed substantial channel sensitivity across much of the visible range from approximately 400 to 700 nm, while their added triple-band optical filter selected bands centered at 457, 530, and 628 nm. Thus, even the camera red channel does not reproduce the conventional red/NIR pulse oximetry pair, and an ordinary RGB camera provides no independently controlled near-infrared channel near 940 nm. Importantly, these spectral sensitivity functions vary among camera models, so the spectral response of one camera should not be treated as representative of all consumer RGB devices [42]. These spectral differences are illustrated in Figure 3.

The main spectral and measurement differences between conventional pulse oximetry and RGB-camera sensing are summarized in Table 3.

Table 3. Quantitative spectral comparison between conventional pulse oximetry and representative RGB-camera sensing.

Property	Conventional Pulse Oximetry	RGB-Camera Sensing
Primary spectral information	Red $\approx$ 660 nm + NIR $\approx$ 940 nm	Broad R, G, B visible responses
Spectral region	Red + near-infrared	Mainly visible RGB
Representative camera response	Not applicable	Broad overlapping sensitivity across approximately 400–700 nm in the example reported in [14]

Table 3. Cont.

Property	Conventional Pulse Oximetry	RGB-Camera Sensing
Example added optical filtering	Not applicable	Triple-band filter centered at 457, 530, 628 nm in the modified system of Gastel and Verkruyse [14]
NIR information near 940 nm	Explicitly measured	Not independently available in ordinary RGB output
Device dependence	Optical wavelengths deliberately selected by device design	RGB spectral sensitivity varies between camera sensors
Main implication for SpO ₂	Two controlled channels support calibrated RoR measurement	RGB ratios depend strongly on camera response, illumination, and calibration

The RGB spectral ranges shown here are representative rather than universal because spectral sensitivity varies among camera sensors and acquisition systems [42].

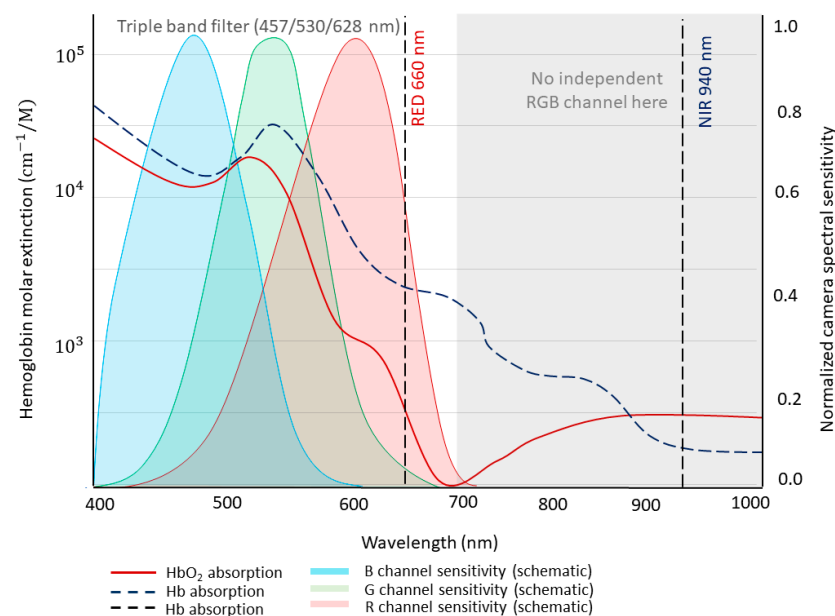


Figure 3. Representative comparison of spectral characteristics between conventional pulse oximetry and RGB-camera sensing. Pulse oximetry uses two narrow bands (red ≈ 660 nm, NIR ≈ 940 nm), whereas RGB channels are broad, overlapping visible-range sensitivities with no independent NIR channel near 940 nm. Curves are schematic and representative; actual camera sensitivity is device-dependent.

5.2. Camera Pipeline Distortion

Consumer cameras are designed to produce visually pleasing images, not to preserve biomedical optical ratios. Auto-exposure, auto-white balance, gamma correction, demosaicing, compression, denoising, and color correction can all change the relative amplitudes and baselines of RGB traces [42,51]. These operations may be harmless for visual appearance, but they can distort the AC/DC relationships that SpO₂ algorithms try to use [15,16]. Because camera models differ in sensor response and processing pipelines, the same physiological signal may therefore be represented differently across devices. At the sensor level, bit depth, read noise, gain, exposure time, saturation, and frame rate stability can further determine how faithfully the small pulsatile RGB variations are captured before any signal-processing or SpO₂ estimation step.

Video encoding and compression introduce an additional source of domain shift. Chroma subsampling, quantization, bitrate reduction, temporal prediction, and compression-related smoothing may modify the subtle color variations that carry rPPG information. A model trained on raw or minimally processed video may therefore not retain the same

performance when applied to compressed webcam, smartphone, telemedicine, or stored video streams. This issue is recognized experimentally in the literature: some RGB-only SpO₂ studies deliberately disabled automatic camera processing or video encoding during acquisition. Wu et al. disabled auto-exposure, autofocus, and auto-white balance [21], while Chan et al. acquired video without auto-tuning or encoding [28].

These acquisition differences should be treated as a domain-generalization problem rather than only as implementation details. Evaluation using multiple camera models does not necessarily demonstrate device-independent performance if the same devices are represented during model development or calibration. Stronger evidence requires testing on unseen cameras, codecs, or processing pipelines without recalibration. RGB-only SpO₂ studies should therefore report whether videos were acquired as raw, linear, or compressed RGB data and should document exposure control, white balance, gain, frame rate, resolution, codec, and compression settings whenever available.

5.3. Skin Optics and Fairness

Skin optics are not a secondary nuisance variable in visible-light SpO₂ estimation. Melanin absorption, scattering, perfusion, surface reflection, temperature, and local vascular response can change the measured color signal [52,53]. These effects are especially important for RGB systems, as visible wavelengths are more strongly influenced by pigmentation than near-infrared wavelengths [54,55]. A method that performs well in one demographic group may fail in another, even when the algorithm and camera are unchanged, particularly when skin-tone distribution and subgroup performance are not reported [29,56].

5.4. Narrow Saturation Range Bias

Narrow SpO₂ distributions can make performance appear stronger than it is. If most samples lie between 96% and 100%, a model can achieve a low average error while learning little about clinically important desaturation. The correct test is not whether the model predicts healthy resting subjects well. The correct test is whether it remains reliable when SpO₂ falls into ranges where clinical decisions matter [57,58]. RGB-only SpO₂ studies should report saturation distribution, low-oxygen saturation sample counts, and range-stratified errors rather than relying only on pooled MAE or RMSE.

6. Taxonomy of RGB-Only Contactless SpO₂ Estimation Methods

A useful taxonomy for RGB-only SpO₂ estimation should not be based only on algorithm type. For SpO₂ estimation, one of the first constraints is the sensing input, since the camera determines what oxygenation-related information can be captured before any signal-processing or learning model is applied [14–16]. In this review, RGB-only methods are therefore organized using five linked dimensions: sensing constraint, signal representation, estimation principle, calibration strategy, and validation claim level. Within the estimation-principle dimension, studies are classified according to the dominant mechanism used to convert RGB-derived information into an SpO₂ estimate. Physical or ratio-based methods use explicit optical or pulsatile relationships such as ratio-of-ratios, regression-based methods learn statistical mappings from engineered RGB or rPPG features, deep learning methods learn the representation and/or SpO₂ mapping directly from video, spatiotemporal, or waveform inputs, and physics-informed or hybrid methods combine learned components with physiological, optical, or measurement-model constraints. When a study contains elements from more than one category, classification is based on the dominant mechanism responsible for the final SpO₂ estimate.

Non-RGB systems such as RGB-NIR, multispectral, and hyperspectral imaging are not treated as core RGB-only categories; they are used only as comparison cases to clarify the spectral limitations of RGB-only sensing [59–62]. Figure 4 summarizes this taxonomy. The main point is that RGB-only methods may differ in model complexity, but all operate under the same basic sensing limitation: broad, overlapping, and device-dependent visible camera channels [15]. The taxonomy therefore separates methods that adapt pulse oximetry logic to RGB traces, methods that learn empirical mappings from handcrafted RGB/rPPG features, methods that learn directly from video or spatiotemporal representations, and hybrid or physics-informed methods that attempt to include optical constraints. Each method group should be interpreted together with its calibration requirement and the level of validation evidence supporting its claim.

1. Sensing input	2. Signal form	3. Estimation method	4. Calibration strategy	Validation claim level
Consumer RGB video broad visible camera channels	Mean RGB traces global ROI averaging	RGB ratio / RoR methods AC/DC ratio logic	Subject-specific per-user mapping	Level 1 Feasibility evidence
Controlled RGB video fixed light/camera, still RGB-only	Patch-level traces local skin regions	Feature-based regression handcrafted RGB/rPPG features	Device / light-specific camera or illumination mapping	Level 2 Controlled lab estimation
Beyond RGB-only Comparison only: RGB-NIR, multispectral, hyperspectral	Spatiotemporal maps video-to-map representation	DL Deep / transformer video, maps, or traces learning	Population-level shared mapping across subjects	Level 3 Subject-independent validation
⚠ Main limitation spectral ambiguity + camera dependence	⚠ Risk appearance/lighting cues	Hybrid / physics-informed optical constraints + learning	⚠ Risk bias and limited transfer	Level 4 Cross-device / cross-dataset validation
		⚠ Core caution models cannot create missing spectra		Level 5 Hypoxemia + clinical agreement

Figure 4. Taxonomy of RGB-only contactless SpO₂ estimation methods. The taxonomy begins with the sensing constraint because all RGB-only methods operate with broad, overlapping, and device-dependent visible camera channels. Methods are then interpreted according to signal representation, estimation principle, calibration strategy, and validation claim level. Non-RGB modalities are included only as comparison cases and are not treated as core RGB-only evidence.

The following subsections describe the main estimation principles within this taxonomy. These categories are not mutually exclusive because some studies combine RGB ratio features, regression models, deep learning, and calibration strategies. Therefore, each method category should be interpreted together with its sensing condition and validation strength.

6.1. Physics-Based RGB Ratio Methods

Physics-based RGB methods usually begin from the pulse oximetry ratio-of-ratios principle [5,7]. They estimate pulsatile and baseline components from two or more RGB channels and then map these features to SpO₂ through calibration [20]. Early camera-based work showed that visible-light video can contain oxygenation-related information under ambient-light conditions, but also emphasized the need for careful calibration and controlled interpretation [63]. The main advantage of these methods is interpretability. If the result fails, the pipeline can be inspected at the level of ROI selection, filtering, AC/DC estimation, channel ratio formation, and calibration.

The limitation is that an RGB ratio is not equivalent to a red/near-infrared pulse oximetry ratio. RGB channels are broad camera responses, not controlled wavelength bands. Therefore, a ratio computed from RGB traces may carry oxygenation-related information under controlled conditions, but its physiological meaning depends strongly on illumination, camera response, skin absorption and scattering characteristics, preprocessing, and calibration. These methods should therefore be interpreted as calibrated RGB estimation methods, not as direct replacements for conventional pulse oximetry.

6.2. Classical Feature-Based Regression Methods

Classical regression methods use handcrafted RGB- or rPPG-derived features as input to estimate SpO_2 . These features may include channel ratios, normalized pulse amplitudes, AC/DC-related features, spectral descriptors, temporal statistics, or combinations of color and pulse-quality features. Regression models are then used to learn the relationship between these extracted features and reference SpO_2 values [21,22].

The strength of these methods is that the feature pipeline is usually easier to inspect than a deep neural network. If the prediction fails, the error can often be traced back to unstable ROI selection, weak pulse amplitude, poor AC/DC estimation, illumination change, or an unsuitable calibration set. This makes classical regression useful for understanding which parts of the RGB signal may contain oxygenation-related information.

These methods remain limited by the information available in the RGB signal. A regression model can map existing RGB features to SpO_2 , but it cannot create missing spectral specificity. If the RGB channels do not contain stable oxygenation-dependent contrast, the model may learn dataset-specific relationships instead of a general physiological mapping. This risk becomes stronger when the dataset has a narrow SpO_2 range, repeated subjects, fixed lighting, or a weak validation split [7,32].

For this reason, classical machine learning methods should not be judged only by MAE or RMSE. Their validity depends on whether the features remain stable across subjects, skin tones, cameras, and lighting conditions, and whether the calibration is tested outside the setting where it was learned. A low error under controlled conditions may support constrained RGB-based estimation, but it does not by itself establish a reliable absolute SpO_2 measurement.

6.3. Deep Learning- and Transformer-Based Methods

Deep learning methods can learn nonlinear mappings from videos, spatiotemporal maps, or extracted rPPG traces to SpO_2 . CNN-based models may exploit spatially distributed facial information, while recurrent and transformer-based models may capture longer temporal dependencies. These are reasonable modeling choices when signal quality varies over time and location [25,26,64]. Deep learning architectures developed for general rPPG, including attention-based [65], spatiotemporal [51], temporal-shift [66], and contrastive-learning [67] models, provide useful methodological background for video-based physiological estimation. However, they should not be treated as direct evidence for RGB-only SpO_2 estimation unless SpO_2 is explicitly estimated and validated.

The interpretation of deep learning performance depends on the validation design. A network can reduce noise, exploit temporal structure, and learn interactions that handcrafted methods miss. It cannot recover a missing narrowband NIR measurement that the camera never captured. Apparent success may therefore reflect physiological oxygenation information, but it may also reflect subject identity, camera response, lighting, skin-tone, perfusion state, or dataset-specific label distributions [29,56].

Explainability tools such as feature-map visualization [24], Grad-CAM [68], SHAP [69], or Integrated Gradients [70] may help identify whether model predictions depend on physiologically plausible facial regions or signal features. Cheng et al. [24] provided a

specific example using hidden-layer feature-map visualization in an EfficientNet-B3 model trained on RGB spatiotemporal maps. After aligning an interpolated feature-map with the input facial frames, they reported greater weighting around the forehead, nose, and cheeks, while the torso and regions without a face received less weight. This spatial pattern is consistent with facial regions commonly used for rPPG extraction. The analysis was limited to interpretable lower-level feature-maps and does not establish that the model learned an oxygenation-specific physiological mechanism. Therefore, explainability methods should complement, rather than replace, subject-independent, cross-device, and physiologically constrained validation.

For deep learning methods, subject-independent validation is essential, given that facial appearance, skin-tone, camera response and lighting conditions can act as identity-related shortcuts that reduce test error without proving oxygenation-sensitive measurement. The main validation risks for deep learning-based RGB-only SpO₂ estimation are summarized in Table 4.

Table 4. Deep learning risk factors in rPPG-based SpO₂ estimation.

Risk	Why It Matters	Recommended Test
Subject leakage	Facial identity can inflate accuracy	Use subject-independent validation
Session leakage	Repeated lighting or camera settings can leak into testing	Use session-independent validation
Normal-range bias	Low MAE can occur without desaturation detection	Report errors by SpO ₂ range
Camera bias	The model may learn sensor response instead of physiology	Use cross-camera or cross-dataset testing
Skin-tone bias	Visible-light signals affected by pigmentation and scattering	Report subgroup performance
Label delay	Oximeter averaging can misalign labels	Analyze time-delay sensitivity

6.4. Hybrid or Physics-Informed RGB Methods

Hybrid and physics-informed methods can be viewed as an emerging direction in RGB-only SpO₂ estimation. The motivation is technically reasonable: purely data-driven models may learn dataset-specific color patterns, while ratio-based RGB methods can be sensitive to illumination, camera response, and skin optics [14,24,25]. A hybrid approach can, in principle, combine optical constraints with learned components, ensuring the model does not treat SpO₂ prediction as an unconstrained regression problem.

For RGB-based systems, however, the term “physics-informed” should be used with caution. A model should not be considered physics-informed merely because it refers to the Beer–Lambert law or uses physiological terminology. The physical component should be connected to measurable or physiologically meaningful quantities in the RGB measurement process, such as hemoglobin absorption and tissue optics [5], pulsatile AC/DC relationships [7], or camera spectral sensitivity and calibration [14,50]. If these relationships are not explicitly incorporated into the model or evaluated experimentally, the physical interpretation may remain post hoc rather than functioning as an actual model constraint.

A related example within remote physiological measurement is PhysMLE, which incorporates prior physiological relationships into a multitask learning framework that includes blood oxygen saturation and evaluates generalization using a multi-source domain-generalization protocol [71]. Although this framework is not a direct physical model of RGB-based SpO₂ formation, it illustrates how physiological prior knowledge can be incorporated into learning while explicitly considering generalization across acquisition domains. For RGB-only SpO₂ estimation, a more strongly physics-informed formulation would additionally connect such learning constraints to the optical measurement process,

including hemoglobin absorption, camera spectral response, AC/DC behavior, tissue optics, and calibration.

The main challenge is that RGB channels do not provide controlled red and near-infrared measurements. Therefore, even a physics-informed RGB model remains limited by the spectral information captured by the camera [14,15]. Such methods may improve robustness by guiding feature learning, reducing implausible predictions, or enforcing calibration-aware behavior, but they still require careful validation across subjects, lighting conditions, cameras, skin tones, and low-SpO₂ ranges [17,29,56].

At present, hybrid and physics-informed RGB-only SpO₂ estimation should be discussed as a developing research direction rather than as established clinical evidence. Its value depends on whether the study shows that the physical assumptions are operational in the model and whether the validation demonstrates improved generalization, not only improved average error in one dataset.

7. Calibration Strategies and Claim Strength

Calibration determines how far an RGB-only SpO₂ method can be interpreted beyond the acquisition condition in which it was developed. Subject-specific calibration may be useful for tracking oxygenation-related changes in one person, but it does not establish population-level SpO₂ estimation. Device-specific or illumination-specific calibration can improve stability, but it also means that the method depends on a controlled camera and lighting environment. Population-level calibration is more scalable, but it requires a sufficiently diverse calibration set across subjects, skin tones, cameras, illumination conditions, and SpO₂ ranges [16,23].

For RGB-only systems, calibration should be treated as a central part of the method rather than a minor implementation detail. Calibration can learn an empirical mapping between RGB-derived features and reference SpO₂, but it cannot recover spectral information that was not captured by the camera [7,15]. Therefore, a method that performs well under fixed lighting, a known camera, or per-subject calibration should be described as constrained RGB-based SpO₂ estimation, not general clinical deployment.

Claim strength should match validation strength. A subject-specific model only supports individual-level or trend-monitoring claims. A model tested with subject-independent validation may support population-level feasibility in similar conditions. Cross-device or cross-dataset validation provides stronger evidence of generalization. Clinical relevance requires low-saturation or hypoxemia-focused testing, agreement analysis, and subgroup evaluation [14,28,32]. Without this evidence, low average error should not be interpreted as proof of clinically reliable SpO₂ measurement. A calibration model trained under one camera, illumination condition, subject group, or SpO₂ distribution should not be assumed to transfer to another condition unless this transfer is directly tested. For RGB-only systems, calibration transfer is a measurement problem, not only a model-generalization problem.

8. Dataset and Ground-Truth Limitations

The most common weakness in contactless SpO₂ papers is not the absence of an advanced model, but weak validation. Datasets dominated by healthy resting subjects can make prediction appear accurate while providing little evidence about hypoxemia. For SpO₂, the distribution of the target variable is not a secondary detail; it determines what kind of clinical claim the study can support.

8.1. SpO₂ Distribution

Every RGB-only contactless SpO₂ study should report the minimum, maximum, mean, standard deviation, histogram, and number of samples below clinically relevant

thresholds such as 95%, 92%, and 90%. Without this information, a low MAE cannot be interpreted properly [31,32]. A model evaluated mostly between 96% and 100% may appear accurate solely because of the narrowness of the test problem. It may still fail when oxygen saturation falls into a range where clinical decisions are made [23,28]. When the test set is dominated by normal SpO₂ values, a model can obtain a low MAE by predicting values near the population mean, even if it has little ability to detect clinically meaningful desaturation [21].

Many public rPPG datasets were originally designed for pulse or general physiological monitoring rather than hypoxemia-focused SpO₂ validation. Datasets such as VIPL-HR [72], PURE [73], and UBFC [44] are useful for studying video-based physiological sensing, but they do not by themselves solve the need for low-SpO₂ coverage, synchronized reference oximetry and subgroup validation.

8.2. Ground-Truth Pulse Oximeter Limitations

The reference pulse oximeter is not an instantaneous or error-free label source. Device model, sampling rate, averaging window, sensor location, motion artifact, and reporting delay all affect the label [1,5]. These issues become more important during desaturation and recovery, when SpO₂ changes over time. If the video window and oximeter label are not temporally aligned, a model may be penalized or rewarded for the wrong physiological state. For desaturation and recovery experiments, the temporal delay and averaging window of the reference pulse oximeter should be reported or analyzed, because a mismatch between the video window and the oximeter label can bias the apparent prediction error [14,28].

8.3. Split Protocol

Frame-level or random window-level splits are weak for video-based physiological sensing, as they can leak subject, session, camera, and lighting information across training and test sets. Subject-independent validation should be considered the minimum acceptable split for population-level claims. Session-independent validation is stronger. Cross-camera or cross-dataset validation provides more meaningful evidence that the model is not simply learning acquisition-specific bias [24,32].

For window-based analysis, overlapping or temporally adjacent windows from the same recording should not be divided across training and test sets, since this can introduce leakage through shared subject identity, illumination, camera settings, motion pattern, and physiological state [32]. Hyperparameter tuning, model selection, normalization statistics, and calibration fitting should also be performed without using test subjects or test sessions. Otherwise, the reported performance may still be optimistic even when the final split appears subject-independent.

8.4. Skin-Tone and Demographic Generalization

Visible-light rPPG is sensitive to skin optics, so demographic information should not be treated as optional. Fitzpatrick skin type or another skin-tone descriptor, age, sex, clinical condition, and perfusion state should be reported when available. If subgroup analysis is not possible, the limitation should be stated explicitly. A method that has not been tested across skin tones should not be described as broadly deployable [29].

For RGB-only SpO₂ estimation, skin-tone reporting is not only a fairness consideration, but also a measurement-validity requirement. Visible-light signals are affected by melanin absorption, scattering, local perfusion, and surface reflection, but a model validated mainly on a narrow skin-tone distribution may show low average error while failing under different pigmentation or perfusion conditions [56]. Therefore, RGB-only SpO₂ studies should report skin-tone distribution, demographic composition, and subgroup performance whenever

possible. If subgroup analysis is not possible, the absence of such an evaluation should be explicitly stated as a limitation [20].

Weak peripheral or facial perfusion can reduce pulsatile amplitude, making AC/DC estimation unstable even when the video appears visually usable [36]. This problem is especially relevant during cold exposure, vasoconstriction, illness, motion, or low-signal-quality conditions. Therefore, studies should report whether low-perfusion cases were included, excluded, or treated through signal-quality control. When subgroup data are available, the reporting items in Table 5 should be presented separately for relevant skin-tone or demographic groups rather than only as pooled overall performance. Figure 5 then places this requirement within the broader validation hierarchy, where subgroup evaluation is part of the evidence needed before RGB-only SpO₂ methods can support broad or clinically interpretable claims.

Table 5. Dataset and validation information required for interpreting contactless SpO₂ studies.

Validation Item	Why It Matters	Reviewer Expectation
SpO ₂ range	Determines whether the model was tested where clinical decisions occur.	Report min, max, mean, distribution, and low-SpO ₂ counts.
Hypoxemia inclusion	Normal-range accuracy does not prove desaturation detection.	Report performance below 95%, 92%, and 90% when available.
Split protocol	Random frame/window splits can leak subject and session information.	Use subject-independent splits at minimum.
Cross-device testing	RGB camera response is device-dependent.	Test on different camera or dataset when possible.
Skin-tone analysis	Visible-light PPG is affected by pigmentation.	Report subgroup performance or acknowledge missing analysis.
Ground-truth	Pulse oximeter type, delay, and averaging affect labels.	Report device, sampling rate, averaging window, and delay correction.
Motion and lighting	Real deployment rarely has perfect conditions.	Evaluate static and challenging conditions separately.

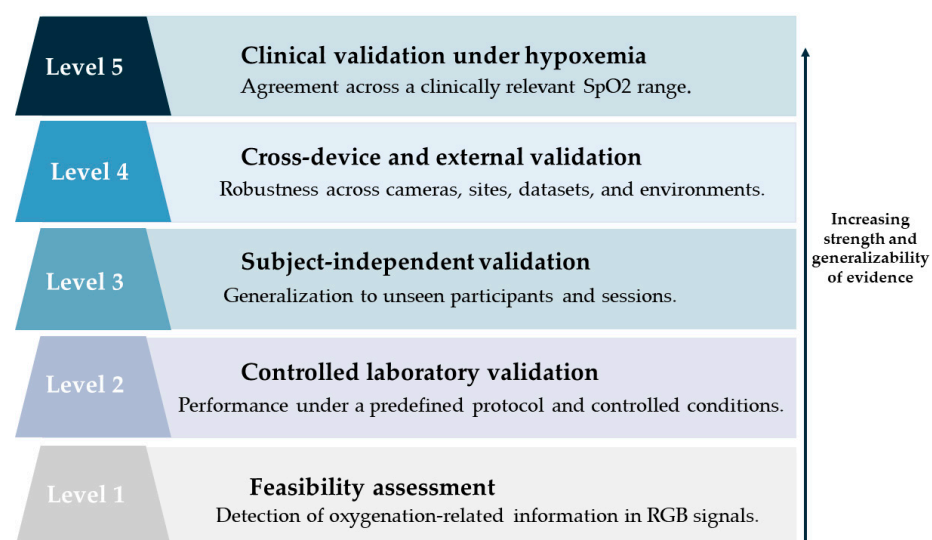


Figure 5. Validation hierarchy for contactless SpO₂ estimation. The five levels represent progressively stronger evidence, ranging from proof-of-concept signal association to clinical agreement across a clinically relevant hypoxemic range. Higher levels require broader validation across participants, sessions, devices, environments, and clinical conditions. Detailed reporting requirements for each level are summarized in Table 5.

Table 5 summarizes the reporting information needed to interpret individual RGB-only SpO₂ studies. Figure 5 then organizes these requirements into a validation hierarchy, linking study design to the strength of the claim being made.

9. Evaluation Metrics and Clinical Validation

Evaluation metrics should match the intended claim. MAE and RMSE are useful summaries, but they are not sufficient for clinical interpretation. A method can have a low average error and still show systematic bias, wide limits of agreement, poor low-saturation performance, or unreliable hypoxemia detection. Correlation is also insufficient because a high correlation can coexist with clinically unacceptable absolute error [30,31]. The commonly reported error and agreement metrics can be written as

$$MAE = \left(\frac{1}{N} \right) \sum_i |y_i - \hat{y}_i| \quad (7)$$

$$RMSE = \sqrt{\left(\frac{1}{N} \right) \sum_i (y_i - \hat{y}_i)^2} \quad (8)$$

$$Bias = \left(\frac{1}{N} \right) \sum_i (\hat{y}_i - y_i) \quad (9)$$

$$LoA = Bias \pm 1.96 SD_{error} \quad (10)$$

$$A_{rms} = \sqrt{\left(\frac{1}{N} \right) \sum_i (\hat{y}_i - y_i)^2} \quad (11)$$

where y_i is the reference oxygen-saturation value, $\hat{y}_i$ is the estimated SpO₂ value, N is the number of samples, and SD_{error} is the standard deviation of the estimation error. Under the displayed definition, A_{rms} is numerically equivalent to RMSE calculated from the same paired measurements. The notation A_{rms} is retained because it is conventionally used in pulse-oximeter performance evaluation against reference arterial oxygen saturation. Metrics should also be reported separately for normal and low-saturation ranges whenever the dataset allows. Table 6 summarizes the main evaluation metrics used in contactless SpO₂ studies and highlights what each metric can and cannot show.

Table 6. Metrics for contactless SpO₂ evaluation.

Metric	What It Shows	What It Hides
MAE	Average absolute error.	Bias, low-SpO ₂ failure, and distribution imbalance.
RMSE	Penalty for large errors.	Whether errors occur in clinically important ranges.
Pearson r	Linear association or trend agreement.	Absolute bias and calibration error.
Bland–Altman analysis	Bias and limits of agreement.	Requires sufficient range and sample diversity.
A_{rms} -like error	Pulse oximetry-style overall error.	Needs clinical context and saturation range reporting.
Sensitivity below threshold	Ability to detect hypoxemia.	Requires enough low-SpO ₂ samples.
Calibration error	Reliability of predicted values.	Rarely reported in current rPPG-SpO ₂ work.

Regulatory pulse oximetry standards and clinical agreement metrics are used in this review as interpretive benchmarks, not as evidence that RGB-only camera-based systems currently satisfy the requirements of clinical pulse oximeters [57,58]. Pearson correlation should not be interpreted as agreement. A method can show high correlation with the reference while still having systematic bias or wide limits of agreement, which would make the absolute SpO₂ estimates clinically unreliable.

10. Critical Comparison of Representative Studies

Representative studies are compared across three linked dimensions: reported quantitative performance, technical design, and validation strength. Numerical accuracy should be interpreted together with the sensing and processing choices of each study and with the evidence supporting the claimed level of generalization. Table 7a–c summarizes the representative quantitative and technical characteristics, while Table 8 compares the validation strength and claim level. The full set of 13 core RGB-only studies is identified in Section 2.1, and the complete study-level quantitative comparison is provided in Supplementary Table S2.

Table 7. (a) Examples illustrating the effect of validation design and SpO₂ range on reported performance. (b) Experimental setup and reference measurement. (c) Signal processing and estimation method.

(a)					
Study	SpO ₂ Range	Validation	MAE/RMSE (%)	Main Point	
Kim et al. [20]	85–100%	Same-dataset fit/evaluation	0.537/0.692	Very low error, but no independent validation	
Wu et al. [21]	Broad low range	LOSO + cross-dataset	4.22–4.45/4.91–5.23	Stronger generalization test; error increases at low SpO ₂	
Cheng et al. [24]	~86–99%	Subject-wise CV	1.274/1.710	MAE rises to 3.077% below 95%	
Chan et al. [28]	~81–99%	LOSO	1.65/1.93 best condition	RMSE varies from 1.93% to 15.66% across lighting	
(b)					
Study	Modality	Input	ROI/Region	Calibration/Validation Strategy	Ground-Truth
Wu et al. [21]	RGB camera	Image-derived color traces	Cheek	LOSO, cross-dataset validation	Fingertip pulse oximeter
van Gastel et al. [14]	RGB camera with added low-cost components	RGB video	Forehead	Hypoxia-lab calibration, robust regression	Finger/ear/nasal pulse oximeters, hypoxia
Nishidate et al. [39]	rPPG-based SpO ₂	RGB video	Forehead	Color checker calibration, white balance correction, empirical calibration	Finger pulse oximeter, respiration transducer, tissue oxygen monitor
Cheng et al. [24]	rPPG-based SpO ₂	Facial RGB video	Forehead, nose, cheeks	5-fold subject cross-validation	VIPL-HR reference SpO ₂
Peng et al. [25]	Video-based rPPG	RGB video	Face and hand skin ROIs	leave-one-participant-out evaluation	CMS50E pulse oximeter
Akamatsu et al. [26]	Video-based rPPG	RGB facial video	facial ROI grid below eyes	five-fold subject-wise cross-validation	OxyTrue A pulse oximeter
(c)					
Study	Preprocessing			Model/Principle	
Wu et al. [21]	CHROM/POS extraction, motion reconstruction, adaptive filtering			Beer–Lambert + RoR, KNN regression	
van Gastel et al. [14]	MOSSE tracking, downsampling, DC normalization, Butterworth bandpass filtering			APBV, RoR (Red/Blue), pulse-spectrum skewness weighting	
Nishidate et al. [39]	RGB-to-XYZ conversion, chromophore estimation, FIR bandpass filtering			Monte Carlo skin model, diffuse reflectance spectroscopy, StO ₂ /SpO ₂ estimation from HbO/HbR	

Table 7. Cont.

Cheng et al. [24]	STMap generation, RGB/YUV/YCrCb conversion, resizing, normalization	Deep learning regression, ResNet-50/DenseNet-121/EfficientNet-B3
Peng et al. [25]	YOLO detection, ROI scaling, FIR filtering, chromatic augmentation, handcrafted RGB statistics	Beer–Lambert + RoR, 3DCNN, CNN-BiLSTM, MFAM attention
Akamatsu et al. [26]	Face alignment, STMap generation, DC/AC filtering, normalization	RoR principle, ResNet18 CNN, DC/AC fusion

Table 8. Validation strength and claim level of representative studies.

Study	Subjects	SpO ₂ Range	Low-SpO ₂ Samples	Split Protocol	Cross-Device?	Metrics	Claim Level
Wu et al. [21]	60 healthy subjects; SSL: 14, NCTUBO: 46	~65–100%	SSL mean: 87.98%; NCTUBO: 1.21% below 90%	LOSO and cross-dataset validation	Yes	MAE, STD of error, RMSE	Level 4
van Gastel et al. [14]	Dataset I: 19 subjects; Dataset II: 10 subjects	70–100%	Six hypoxia levels (94%, 90%, 85%, 80%, 75%, 70%)	Dataset I for calibration; Dataset II for realistic validation	No	Bias, absolute error, pulse-spectrum skewness	Level 4
Nishidate et al. [39]	13 healthy volunteers	87–100%	Hypoxia induced to ~87% using 11% oxygen inhalation	No standard train/test split; empirical model from Monte Carlo simulations	No	RMSE, Pearson correlation, Bland–Altman analysis	Level 2
Cheng et al. [24]	VIPL-HR: 107 subjects	~88–100%	Normal ($\geq 95\%$) vs. abnormal ($< 95\%$) groups	Five-fold subject cross-validation	Yes	MAE, RMSE	Level 3
Peng et al. [25]	MVPD: 13 volunteers; FaceForensics++: 1000 videos; UBFC: 42 videos	93–100%	Breath-holding tests within 93–100%	Contrastive pretraining + supervised fine-tuning + leave-one-out testing	No	MAE, RMSE, Pearson correlation	Level 3
Akamatsu et al. [26]	50 subjects	85–100%	Breath-holding cycles (30 s hold + 30 s rest)	Five-fold subject-wise cross-validation	No	MAE, RMSE, Pearson correlation coefficient	Level 3

10.1. Quantitative Cross-Study Comparison

Reported MAE and RMSE values across RGB-only SpO₂ studies should not be interpreted as directly comparable performance measures because the underlying evaluation conditions differ substantially. The included studies vary in SpO₂ range, calibration strategy, subject independence, camera hardware, acquisition conditions, and validation design. For example, Kim et al. [20] reported a MAE of 0.537% over an 85–100% range, but the empirical calibration and evaluation appear to have been performed on the same dataset. In contrast, Wu et al. [21] reported higher errors under subject-independent and cross-dataset evaluation, with MAE increasing from 3.39% at 95–100% SpO₂ to 5.57% at 80–85%. Similarly, Cheng et al. [24] reported an overall MAE of 1.274%, but the error increased to 3.077% for samples below 95% SpO₂. These examples show that a low reported overall error can reflect restricted target range or less stringent validation rather than stronger measurement validity.

For this reason, a conventional forest plot or pooled meta-analytic effect size was not calculated. The reported errors do not represent a common estimate under comparable experimental conditions, and statistical pooling could therefore imply a level of comparability that is not supported by the study designs. Instead, we performed a structured

quantitative synthesis that jointly considers reported error, SpO₂ coverage, calibration independence, validation design, low-SpO₂ performance, and external or device-level testing. This approach also highlights that acquisition conditions can strongly influence apparent performance; for example, Chan et al. [28] reported RMSE values ranging from 1.93% to 15.66% across lighting conditions using the same general measurement framework. Table 7a summarizes representative examples showing how reported performance changes with SpO₂ range, validation design, and acquisition conditions.

The complete study-level quantitative comparison is provided in Supplementary Table S2.

10.2. Technical Design and Signal Processing

To provide technical context for the reported performance, Table 7b,c summarizes the acquisition, reference measurement, preprocessing, and estimation choices of representative RGB-only studies.

10.3. Validation Strength and Claim Level

For interpreting validation strength, this review uses five claim levels. Level 1 indicates feasibility evidence under limited or controlled conditions. Level 2 indicates constrained laboratory estimation with limited generalization evidence. Level 3 indicates subject-independent validation under similar acquisition conditions. Level 4 indicates cross-condition, cross-device, or cross-dataset generalization evidence. Level 5 indicates clinically relevant validation with low-SpO₂ or hypoxemia coverage, agreement analysis, and evidence supporting absolute SpO₂ interpretation. These levels are assigned according to the reported validation evidence, not according to algorithm complexity.

The comparison should be interpreted conservatively. The assigned claim level is not a judgment of algorithmic complexity alone, but reflects the strength of the validation evidence supporting clinical interpretation. When a study satisfied some but not all criteria for a higher level, the lower conservative level was assigned. A technically advanced model evaluated on a narrow, subject-overlapping, or weakly validated dataset should not be treated as high-level clinical evidence. Conversely, a simpler calibrated method tested across a wider SpO₂ range, with stronger reference measurement and subject-independent validation, may provide stronger measurement evidence.

11. Clinical Absolute SpO₂ vs. Relative Oxygenation Trend Monitoring

Many studies have discussed two different applications as equivalent. The evidence required for these applications is not the same [11,12]. Clinical absolute SpO₂ estimation means the system predicts an actual medical SpO₂ value as a pulse oximeter. In this case, an error near the hypoxemic range is seen as clinically important, as it may change whether a subject is considered normal or at risk. Such claims require low bias, narrow limits of agreement, and reliable performance during hypoxemia [31,57]. Robustness across subjects, devices, lighting conditions, and skin-tone groups is also necessary.

Relative oxygenation trend monitoring is a more limited but more realistic target for RGB-only rPPG. The system may detect whether an oxygenation-related signal is decreasing, recovering, or becoming unstable relative to a subject-specific baseline; it does not estimate the exact SpO₂ value. This application also requires validation across sessions and acquisition conditions, but it does not require the same level of absolute agreement as clinical pulse oximetry [14–16].

RGB-camera-based methods should not be described as fully reliable clinical pulse oximeters, unless a study provides strong calibration, subject-independent validation, low-saturation testing, and clinical agreement analysis. This distinction protects the field

from overclaiming while still giving RGB-only research a defensible role in screening, monitoring, or research settings under controlled conditions.

12. Recommended Reporting Checklist for RGB-Only SpO₂ Studies

A minimum reporting checklist can improve reproducibility and reduce overinterpretation of RGB-only contactless SpO₂ results. The checklist in Table 9 is not a regulatory standard; it is a practical reporting guide for studies that aim to make physiologically meaningful claims using RGB-derived oxygen saturation estimates.

Table 9. Minimum reporting checklist for RGB-only contactless SpO₂ studies.

Reporting Item	Minimum Information That Should Be Reported	Why It Matters
Camera system	Type, frame rate, resolution, exposure, color format	RGB values depend on camera settings
Illumination	Light source, ambient/controlled setting, lighting changes	Light spectrum strongly affects RGB signals
ROI selection	Body region, skin segmentation, tracking, rejected frames	ROI changes affect AC/DC ratios
RGB trace extraction	Averaging method, patch method, channel handling	Defines the input signal
Preprocessing	Filtering, normalization, resampling, window length	Can change pulse amplitude and baseline
AC/DC features	AC/DC definition, channels, amplitude method, features	Different definitions change SpO ₂ estimates
Calibration method	Calibration type, dataset, equation/model	Determines claim strength
Reference SpO ₂	Reference device, sampling, averaging, synchronization	Poor alignment can bias error
Validation	Subject- and session-independent testing; unseen-device or external-dataset evaluation without test-domain recalibration	Weak splits inflate accuracy
Metrics	Bias, limits of agreement, pulse oximetry-style error, saturation-stratified errors, and hypoxemia detection performance; MAE/RMSE as supplementary summaries.	Average error alone is insufficient
Reproducibility	Code/data, parameters, model details, statistics.	Enables independent checking
Claim level	Feasibility, controlled, generalizable, or clinical claim	Claims must match evidence

For a future RGB-only system to support a clinical reliability claim, reporting completeness alone is not sufficient. To provide a more objective benchmark for future RGB-only SpO₂ systems, Table 10 summarizes the quantitative and explicit validation targets anchored to established pulse oximetry guidance and standards. The numerical values are derived from the final FDA pulse-oximeter guidance, while the subject-, session-, calibration-, and cross-device requirements adapt these validation principles to RGB-based machine learning systems.

Table 10. Standard-anchored quantitative readiness targets for future RGB-only SpO₂ systems.

Requirement	Quantitative Target	Basis/Standard
Clinical dataset size	≥10 subjects and ≥200 paired RGB-SpO ₂ /arterial SaO ₂ observations	FDA final guidance [74]
Low-SpO ₂ coverage	Cover 70–100% SaO ₂ ; report results for 70–80%, 80–90%, and 90–100% and report the number and percentage of samples <90% and <85%	FDA final guidance [74]
Subject- and session-independent testing	Required: Yes	RGB/ML validation requirement
Calibration independence	Required: Yes	Clinical-validation principle
Cross-device/domain holdout	Required: Yes	RGB-specific external-validation requirement
Agreement analysis	Report mean bias, Bland–Altman analysis, and upper and lower 95% limits of agreement	FDA final guidance [74]; CLSI [75]
Overall accuracy	Reflectance benchmark: $A_{rms} \leq 3.5\%$ over 70–100% SaO ₂	FDA final guidance [74]
Skin-pigmentation representation	≥2 darkly pigmented subjects or 15% of the sample, whichever is larger	FDA final guidance [74]

13. Future Directions

Future research on RGB-based SpO₂ estimation needs to move from small proof-of-concept results toward scientifically reliable and clinically trustworthy evaluation. Meaningful progress in this field can be achieved through subject-independent, session-independent and cross-device testing. In addition, datasets that include a wide range of SpO₂, hypoxemia samples, skin-tone variation and synchronized reference oximetry are essential for advancement of this field [52,53,55,56,76–79].

Physics-based or uncertainty-aware models may only help if the physical assumptions are linked to measurable RGB features [71,80]. RGB-NIR or multispectral systems may provide stronger wavelength specificity, but they should be treated as beyond-RGB directions rather than direct evidence for RGB-only performance [33–35]. The most urgent technical directions are not only larger neural networks, but standardized acquisition protocols, raw or linear RGB access, cross-camera calibration transfer, low-SpO₂ data collection, subgroup evaluation, and uncertainty estimation that indicates when RGB-only predictions should not be trusted.

14. Conclusions

In this paper, we present a critical review of RGB-only SpO₂ estimation. Even though contactless SpO₂ estimation has a promising future, it is also easy to overstate its capabilities. A consumer-level RGB camera does not provide the controlled spectral information that a clinical pulse oximeter does; this is the central limitation of RGB-only SpO₂ estimation. In our study, we observed physics-based methods that make the optical assumptions easier to inspect, while machine learning and deep learning methods can improve empirical prediction under specific conditions. Existing evidence does not establish a device-invariant and acquisition-independent mapping from broad RGB measurements to absolute SpO₂. Learned models may exploit oxygenation-sensitive information present in RGB video, but their calibration and transferability remain dependent on the camera, illumination, participant population, and validation design. This is why researchers should interpret RGB-based oxygen saturation as a constrained empirical measurement problem. Under controlled conditions, an rPPG signal extracted from an RGB-only camera may be useful for

constrained estimation and for monitoring oxygenation trend. However, clinically reliable SpO₂ estimation requires stronger validation, including subject-independent testing, cross-condition evaluation, agreement analysis, and low-SpO₂ assessment. Future progress should depend less on model complexity alone and more on transparent calibration, robust validation, and claim levels that match the available evidence. Therefore, current RGB-only SpO₂ methods should be described as constrained empirical estimators unless they are validated across low-SpO₂ ranges, subjects, devices, lighting conditions, and relevant demographic subgroups.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/electronics15153477/s1>. Table S1, Structured evidence appraisal of the 13 core RGB-only SpO₂ studies; Table S2, Quantitative performance comparison of the 13 core RGB-only SpO₂ studies.

Author Contributions: Conceptualization, S.I.M.; methodology, S.I.M.; investigation, S.I.M.; formal analysis, S.I.M.; data curation, S.I.M.; writing—original draft preparation, S.I.M. and N.K.K.; writing—review and editing, S.I.M. and N.K.K.; supervision, N.K.K.; project administration, N.K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Regional Innovation System & Education (RISE) program through the Gyeongbuk RISE CENTER, funded by the Ministry of Education (MOE) and Gyeongsangbuk-do, Republic of Korea. (2026-RISE-15-115).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

Abbreviations

The following abbreviations are used in this manuscript:

APBV	Adaptive pulse blood volume method
CHROM	Chrominance-based rPPG method
POS	Plane-orthogonal-to-skin method
CNN	Convolutional neural network
FIR	Finite impulse response
Hb	Deoxygenated hemoglobin
HbO	Oxygenated hemoglobin
HbR	Reduced hemoglobin
HR-rPPG	Heart-rate remote photoplethysmography
ICA	Independent component analysis
PCA	Principal component analysis
KNN	k-nearest neighbors
LOSO	Leave-one-subject-out
MAE	Mean absolute error
RMSE	Root Mean Squared Error
NIR	Near-Infrared
RoR	Ratio-of-Ratios
PBV	Blood volume pulse
STD	Standard deviation
StO ₂	Tissue oxygen saturation

References

- Nitzan, M.; Romem, A.; Koppel, R. Pulse Oximetry: Fundamentals and Technology Update. *Med. Devices Evid. Res.* **2014**, *7*, 231–239. [[CrossRef](#)] [[PubMed](#)]
- Bagha, S.; Shaw, L. A Real Time Analysis of PPG Signal for Measurement of SpO₂ and Pulse Rate. *Int. J. Comput. Appl.* **2011**, *36*, 45–50.
- Aoyagi, T. Pulse Oximetry: Its Invention, Theory, and Future. *J. Anesth.* **2003**, *17*, 259–266. [[CrossRef](#)] [[PubMed](#)]
- Severinghaus, J.W. Discovery of Pulse Oximetry. *Anesth. Analg.* **2007**, *105*, S1–S4. [[CrossRef](#)] [[PubMed](#)]
- Mannheimer, P.D. The Light–Tissue Interaction of Pulse Oximetry. *Anesth. Analg.* **2007**, *105*, S10–S17. [[CrossRef](#)] [[PubMed](#)]
- Zijlstra, W.G.; Buursma, A.; Meeuwse-van der Roest, W.P. Absorption Spectra of Human Fetal and Adult Oxyhemoglobin, de-Oxyhemoglobin, Carboxyhemoglobin, and Methemoglobin. *Clin. Chem.* **1991**, *37*, 1633–1638. [[CrossRef](#)]
- Blaney, G.; Sassaroli, A.; Fantini, S. Critical Analysis of the Relationship between Arterial Saturation and the Ratio-of-Ratios Used in Pulse Oximetry. *J. Biomed. Opt.* **2024**, *29*, S33313. [[CrossRef](#)] [[PubMed](#)]
- Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote Plethysmographic Imaging Using Ambient Light. *Opt. Express* **2008**, *16*, 21434–21445. [[CrossRef](#)] [[PubMed](#)]
- Wang, W.; Den Brinker, A.C.; Stuijk, S.; De Haan, G. Algorithmic Principles of Remote PPG. *IEEE Trans. Biomed. Eng.* **2016**, *64*, 1479–1491. [[PubMed](#)]
- Pirzada, P.; Wilde, A.; Harris-Birtill, D. Remote Photoplethysmography for Heart Rate and Blood Oxygenation Measurement: A Review. *IEEE Sens. J.* **2024**, *24*, 23436–23453. [[CrossRef](#)]
- Saikevičius, L.; Raudonis, V.; Dervinis, G.; Baranauskas, V. Non-Contact Vision-Based Techniques of Vital Sign Monitoring: Systematic Review. *Sensors* **2024**, *24*, 3963. [[CrossRef](#)] [[PubMed](#)]
- Hassanpour, A.; Yang, B. Contactless Vital Sign Monitoring: A Review Towards Multi-Modal Multi-Task Approaches. *Sensors* **2025**, *25*, 4792. [[CrossRef](#)] [[PubMed](#)]
- Sakib, S.; Hasan, Z.; Roy, N. A State-of-The-Art Survey of Remote Photoplethysmography for Contactless Health Parameters Sensing. *WIREs Data Min. Knowl. Discov.* **2025**, *15*, e70039. [[CrossRef](#)]
- van Gastel, M.; Verkruysse, W. Contactless SpO₂ with an RGB Camera: Experimental Proof of Calibrated SpO₂. *Biomed. Opt. Express* **2022**, *13*, 6791–6802. [[CrossRef](#)] [[PubMed](#)]
- Guazzi, A.R.; Villarroel, M.; Jorge, J.; Daly, J.; Frise, M.C.; Robbins, P.A.; Tarassenko, L. Non-Contact Measurement of Oxygen Saturation with an RGB Camera. *Biomed. Opt. Express* **2015**, *6*, 3320–3338. [[CrossRef](#)] [[PubMed](#)]
- Wei, B.; Wu, X.; Zhang, C.; Lv, Z. Analysis and Improvement of Non-Contact SpO₂ Extraction Using an RGB Webcam. *Biomed. Opt. Express* **2021**, *12*, 5227–5245. [[CrossRef](#)] [[PubMed](#)]
- Dasari, A.; Prakash, S.K.A.; Jeni, L.A.; Tucker, C.S. Evaluation of Biases in Remote Photoplethysmography Methods. *npj Digit. Med.* **2021**, *4*, 91. [[CrossRef](#)] [[PubMed](#)]
- Jacques, S.L. Optical Properties of Biological Tissues: A Review. *Phys. Med. Biol.* **2013**, *58*, R37. [[CrossRef](#)] [[PubMed](#)]
- de Haan, G.; Jeanne, V. Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [[CrossRef](#)] [[PubMed](#)]
- Kim, N.H.; Yu, S.-G.; Kim, S.-E.; Lee, E.C. Non-Contact Oxygen Saturation Measurement Using YCgCr Color Space with an RGB Camera. *Sensors* **2021**, *21*, 6120. [[CrossRef](#)] [[PubMed](#)]
- Wu, B.-J.; Wu, B.-F.; Dong, Y.-C.; Lin, H.-C.; Li, P.-H. Peripheral Oxygen Saturation Measurement Using an RGB Camera. *IEEE Sens. J.* **2023**, *23*, 26551–26563. [[CrossRef](#)]
- Al-Naji, A.; Khalid, G.A.; Mahdi, J.F.; Chahl, J. Non-Contact SpO₂ Prediction System Based on a Digital Camera. *Appl. Sci.* **2021**, *11*, 4255. [[CrossRef](#)]
- Shao, Q.; Zhu, L.; Ahmed, M.; Vatanparvar, K.; Gwak, M.; Rashid, N.; Bae, J.; Kuang, J.; Gao, A. Normalization Is All You Need: Robust Full-Range Contactless SpO₂ Estimation Across Users. In Proceedings of the ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 1646–1650.
- Cheng, C.-H.; Yuen, Z.; Chen, S.; Wong, K.-L.; Chin, J.-W.; Chan, T.-T.; So, R.H.Y. Contactless Blood Oxygen Saturation Estimation from Facial Videos Using Deep Learning. *Bioengineering* **2024**, *11*, 251. [[CrossRef](#)] [[PubMed](#)]
- Peng, J.; Su, W.; Chen, H.; Sun, J.; Tian, Z. CL-SPO2Net: Contrastive Learning Spatiotemporal Attention Network for Non-Contact Video-Based SpO₂ Estimation. *Bioengineering* **2024**, *11*, 113. [[CrossRef](#)] [[PubMed](#)]
- Akamatsu, Y.; Onishi, Y.; Imaoka, H. Blood Oxygen Saturation Estimation from Facial Video Via DC and AC Components of Spatio-Temporal Map. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
- Chowdhury, M.H.; Reaz, M.B.I.; Ali, S.H.M.; Khan, M.S.; Chowdhury, M.E.H. ROSE-Net: Leveraging Remote Photoplethysmography to Estimate Oxygen Saturation Using Deep Learning. *Biomed. Signal Process. Control* **2025**, *100*, 107105. [[CrossRef](#)]

28. Chan, M.; Zhu, L.; Vatanparvar, K.; Gwak, M.; Kuang, J.; Gao, A. Estimating SpO₂ with Deep Oxygen Desaturations from Facial Video Under Various Lighting Conditions: A Feasibility Study. 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.
29. Nowara, E.M.; McDuff, D.; Veeraraghavan, A. A Meta-Analysis of the Impact of Skin Type and Gender on Non-Contact Photoplethysmography Measurements. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1148–1155.
30. Bland, J.M.; Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1986**, *327*, 307–310. [[CrossRef](#)]
31. Gerke, O. Reporting Standards for a Bland–Altman Agreement Analysis: A Review of Methodological Reviews. *Diagnostics* **2020**, *10*, 334. [[CrossRef](#)] [[PubMed](#)]
32. Wolff, R.F.; Moons, K.G.M.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S.; for the PROBAST Group. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **2019**, *170*, 51–58. [[CrossRef](#)] [[PubMed](#)]
33. Shao, D.; Liu, C.; Tsow, F.; Yang, Y.; Du, Z.; Iriya, R.; Yu, H.; Tao, N. Noncontact Monitoring of Blood Oxygen Saturation Using Camera and Dual-Wavelength Imaging System. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1091–1098. [[CrossRef](#)] [[PubMed](#)]
34. Shen, C.-H.; Chen, W.-L.; Wu, J.-J. Research on Multiple Spectral Ranges with Deep Learning for SpO₂ Measurement. *Sensors* **2022**, *22*, 328. [[CrossRef](#)] [[PubMed](#)]
35. Arimoto, H. Measurement of 2-D SpO₂ Distribution in Skin Tissue by Multispectral Imaging with Depth Selectivity Control. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; pp. 1968–1971.
36. Arimoto, H. Multispectral Polarization Imaging for Observing Blood Oxygen Saturation in Skin Tissue. *Appl. Spectrosc.* **2006**, *60*, 459–464. [[CrossRef](#)] [[PubMed](#)]
37. Wu, J. Hyperspectral Imaging for Non-Invasive Blood Oxygen Saturation Assessment. *Photodiagnosis Photodyn. Ther.* **2024**, *45*, 104003. [[CrossRef](#)] [[PubMed](#)]
38. Wuyart, A.; Abensur Vuillaume, L.; Maaoui, C.; Bousefsaf, F. A Survey of Blood Oxygen Saturation Assessment from Video. *Biomed. Signal Process. Control* **2025**, *110*, 108069. [[CrossRef](#)]
39. Nishidate, I.; Yasui, R.; Nagao, N.; Suzuki, H.; Takara, Y.; Ohashi, K.; Ando, F.; Noro, N.; Kokubo, Y. RGB Camera-Based Simultaneous Measurements of Percutaneous Arterial Oxygen Saturation, Tissue Oxygen Saturation, Pulse Rate, and Respiratory Rate. *Front. Physiol.* **2022**, *13*, 933397. [[CrossRef](#)] [[PubMed](#)]
40. Lopez, S. Pulse Oximeter Fundamentals and Design. *Free. Semicond.* **2012**, *23*, 1–39.
41. Webster, J.G. *Design of Pulse Oximeters*; CRC Press: Boca Raton, FL, USA, 1997; ISBN 978-1-4200-5079-0.
42. Jiang, J.; Liu, D.; Gu, J.; Süsstrunk, S. What Is the Space of Spectral Sensitivity Functions for Digital Color Cameras? In Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), Clearwater Beach, FL, USA, 15–17 January 2013; pp. 168–179.
43. Li, X.; Chen, J.; Zhao, G.; Pietikainen, M. Remote Heart Rate Measurement from Face Videos under Realistic Situations. In Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, 23–28 June 2014; pp. 4264–4271.
44. Bobbia, S.; Macwan, R.; Benezech, Y.; Mansouri, A.; Dubois, J. Unsupervised Skin Tissue Segmentation for Remote Photoplethysmography. *Pattern Recognit. Lett.* **2019**, *124*, 82–90. [[CrossRef](#)]
45. Kumar, M.; Veeraraghavan, A.; Sabharwal, A. DistancePPG: Robust Non-Contact Vital Signs Monitoring Using a Camera. *Biomed. Opt. Express* **2015**, *6*, 1565–1588. [[CrossRef](#)] [[PubMed](#)]
46. Tulyakov, S.; Alameda-Pineda, X.; Ricci, E.; Yin, L.; Cohn, J.F.; Sebe, N. Self-Adaptive Matrix Completion for Heart Rate Estimation from Face Videos under Realistic Conditions. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2396–2404.
47. Poh, M.-Z.; McDuff, D.J.; Picard, R.W. Non-Contact, Automated Cardiac Pulse Measurements Using Video Imaging and Blind Source Separation. *Opt. Express* **2010**, *18*, 10762–10774. [[CrossRef](#)] [[PubMed](#)]
48. Johnston, W.S. Development of a Signal Processing Library for Extraction of SpO₂, HR, HRV, and RR from Photoplethysmographic Waveforms. Ph.D. Thesis, Worcester Polytechnic Institute, Worcester, MA, USA, 2006.
49. Poh, M.-Z.; McDuff, D.J.; Picard, R.W. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 7–11. [[CrossRef](#)] [[PubMed](#)]
50. Ye, Y.; Pan, L.; Yu, D.; Gu, D.; Lu, H.; Wang, W. Notch RGB-Camera Based SpO₂ Estimation: A Clinical Trial in Neonatal Intensive Care Unit. *Biomed. Opt. Express* **2024**, *15*, 428–445. [[CrossRef](#)] [[PubMed](#)]
51. Yu, Z.; Peng, W.; Li, X.; Hong, X.; Zhao, G. Remote Heart Rate Measurement from Highly Compressed Facial Videos: An End-to-End Deep Learning Solution With Video Enhancement. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 151–160.

52. Bickler, P.E.; Feiner, J.R.; Severinghaus, J.W. Effects of Skin Pigmentation on Pulse Oximeter Accuracy at Low Saturation. *Anesthesiology* **2005**, *102*, 715–719. [[CrossRef](#)] [[PubMed](#)]
53. Feiner, J.R.; Severinghaus, J.W.; Bickler, P.E. Dark Skin Decreases the Accuracy of Pulse Oximeters at Low Oxygen Saturation: The Effects of Oximeter Probe Type and Gender. *Anesth. Analg.* **2007**, *105*, S18–S23. [[CrossRef](#)] [[PubMed](#)]
54. Cabanas, A.M.; Fuentes-Guajardo, M.; Latorre, K.; León, D.; Martín-Escudero, P. Skin Pigmentation Influence on Pulse Oximetry Accuracy: A Systematic Review and Bibliometric Analysis. *Sensors* **2022**, *22*, 3402. [[CrossRef](#)] [[PubMed](#)]
55. Shi, C.; Goodall, M.; Dumville, J.; Hill, J.; Norman, G.; Hamer, O.; Clegg, A.; Watkins, C.L.; Georgiou, G.; Hodgkinson, A.; et al. The Accuracy of Pulse Oximetry in Measuring Oxygen Saturation by Levels of Skin Pigmentation: A Systematic Review and Meta-Analysis. *BMC Med.* **2022**, *20*, 267. [[CrossRef](#)] [[PubMed](#)]
56. Bondarenko, M.; Menon, C.; Elgendi, M. Demographic Bias in Public Remote Photoplethysmography Datasets. *npj Digit. Med.* **2025**, *8*, 593. [[CrossRef](#)] [[PubMed](#)]
57. ISO 80601-2-61:2026; Medical Electrical Equipment—Part 2-61: Particular Requirements for Basic Safety and Essential Performance of Pulse Oximeter Equipment. International Organization for Standardization: Geneva, Switzerland, 2026.
58. Clinical and Laboratory Standards Institute. *Pulse Oximetry; Approved Guideline*; Clinical and Laboratory Standards Institute: Wayne, PA, USA, 2005.
59. Khoobehi, B.; Beach, J.M.; Kawano, H. Hyperspectral Imaging for Measurement of Oxygen Saturation in the Optic Nerve Head. *Investig. Ophthalmol. Vis. Sci.* **2004**, *45*, 1464–1472. [[CrossRef](#)] [[PubMed](#)]
60. Venkat, S.; Arsath, P.S.M.T.P.S.; Alex, A.; Preejith, S.P.; Balamugesh; Christopher, D.J.; Joseph, J.; Sivaprakasam, M. Machine Learning Based SpO₂ Computation Using Reflectance Pulse Oximetry. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 482–485.
61. Ontiveros, R.C.; Elgendi, M.; Missale, G.; Menon, C. Evaluating RGB Channels in Remote Photoplethysmography: A Comparative Study with Contact-Based PPG. *Front. Physiol.* **2023**, *14*, 1296277. [[CrossRef](#)] [[PubMed](#)]
62. Linek, M.; Schrader, I.; Strauß, K.; Volgger, V.; Rühm, A.; Sroka, R.; Aumiller, M. A Comparison between Hyperspectral Imaging and RGB-Based Tissue Perfusion Assessment of Human Skin Using Convolutional Neural Networks. *Comput. Biol. Med.* **2026**, *200*, 111363. [[CrossRef](#)] [[PubMed](#)]
63. Bal, U. Non-Contact Estimation of Heart Rate and Oxygen Saturation Using Ambient Light. *Biomed. Opt. Express* **2015**, *6*, 86–97. [[CrossRef](#)] [[PubMed](#)]
64. Huo, C.; Yin, P.; Fu, B. MultiPhys: Heterogeneous Fusion of Mamba and Transformer for Video-Based Multi-Task Physiological Measurement. *Sensors* **2025**, *25*, 100. [[CrossRef](#)] [[PubMed](#)]
65. Chen, W.; McDuff, D. DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks. In *Computer Vision—ECCV 2018*; Springer: Cham, Switzerland, 2018; pp. 349–365.
66. Liu, X.; Fromm, J.; Patel, S.; McDuff, D. Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 19400–19411.
67. Gideon, J.; Stent, S. The Way to My Heart Is Through Contrastive Learning: Remote Photoplethysmography From Unlabelled Video. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 3995–4004.
68. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2019**, *128*, 336–359. [[CrossRef](#)]
69. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
70. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic Attribution for Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning*; PMLR: New York, NY, USA, 2017; pp. 3319–3328.
71. Wang, J.; Lu, H.; Wang, A.; Yang, X.; Chen, Y.; He, D.; Wu, K. PhysMLE: Generalizable and Priors-Inclusive Multi-Task Remote Physiological Measurement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 4908–4925. [[CrossRef](#)] [[PubMed](#)]
72. Niu, X.; Han, H.; Shan, S.; Chen, X. VIPL-HR: A Multi-Modal Database for Pulse Estimation from Less-Constrained Face Video. In *Proceedings of the Computer Vision—ACCV 2018*; Jawahar, C.V., Li, H., Mori, G., Schindler, K., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 562–576.
73. Stricker, R.; Müller, S.; Gross, H.-M. Non-Contact Video-Based Pulse Rate Measurement on a Mobile Service Robot. In Proceedings of the The 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; pp. 1056–1062.
74. U.S. Food and Drug Administration. *Pulse Oximeters—Premarket Notification Submissions [510(k)s]: Guidance for Industry and Food and Drug Administration Staff*; U.S. Food and Drug Administration: Silver Spring, MD, USA, 2013.
75. CLSI EP09; Measurement Procedure Comparison and Bias Estimation Using Patient Sample. Clinical and Laboratory Standards Institute: Malvern, PA, USA, 2018.

76. Valbuena, V.S.M.; Seelye, S.; Sjoding, M.W.; Valley, T.S.; Dickson, R.P.; Gay, S.E.; Claar, D.; Prescott, H.C.; Iwashyna, T.J. Racial Bias and Reproducibility in Pulse Oximetry among Medical and Surgical Inpatients in General Care in the Veterans Health Administration 2013–2019: Multicenter, Retrospective Cohort Study. *BMJ* **2022**, *378*, e069775. [[CrossRef](#)] [[PubMed](#)]
77. Fawzy, A.; Wu, T.D.; Wang, K.; Robinson, M.L.; Farha, J.; Bradke, A.; Golden, S.H.; Xu, Y.; Garibaldi, B.T. Racial and Ethnic Discrepancy in Pulse Oximetry and Delayed Identification of Treatment Eligibility Among Patients With COVID-19. *JAMA Intern. Med.* **2022**, *182*, 730–738. [[CrossRef](#)] [[PubMed](#)]
78. Parr, N.J.; Beech, E.H.; Young, S.; Valley, T.S. Racial and Ethnic Disparities in Occult Hypoxemia Prevalence and Clinical Outcomes Among Hospitalized Patients: A Systematic Review and Meta-Analysis. *J. Gen. Intern. Med.* **2024**, *39*, 2543–2553. [[CrossRef](#)] [[PubMed](#)]
79. Sjoding, M.W.; Dickson, R.P.; Iwashyna, T.J.; Gay, S.E.; Valley, T.S. Racial Bias in Pulse Oximetry Measurement. *N. Engl. J. Med.* **2020**, *383*, 2477–2478. [[CrossRef](#)] [[PubMed](#)]
80. Song, R.; Wang, H.; Xia, H.; Cheng, J.; Li, C.; Chen, X. Uncertainty Quantification for Deep Learning-Based Remote Photoplethysmography. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5027812. [[CrossRef](#)]

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