Comparative Analysis of Chrominance-Based Techniques for Robust Remote Photoplethysmography

Biomedical Signal Processing Laboratory

Anna Simeone
anna.simeone@mail.polimi.it
245744

I. Introduction

Remote photoplethysmography (rPPG) is a non-contact technique that estimates physiological parameters such as pulse rate (PR) by analyzing subtle skin color changes captured in video recordings. These fluctuations are driven by variations in blood volume during the cardiac cycle, which modulate the skin's optical absorption properties. Compared to contact-based PPG, rPPG offers a convenient and unobtrusive alternative for applications such as telemedicine, fitness monitoring, and surveillance.

Despite its advantages, rPPG remains highly susceptible to motion artifacts and ambient lighting changes. Traditional blind source separation (BSS) techniques—most notably Independent Component Analysis (ICA) and Principal Component Analysis (PCA)—have been proposed to extract the pulsatile signal from raw RGB data by assuming that physiological and non-physiological components are separable based on statistical properties [7].

ICA relies on the assumption of statistical independence between sources, while PCA assumes that the most informative signal corresponds to the direction of highest variance. However, in practical scenarios with motion and dynamic lighting, these assumptions often break down: motion tends to introduce large, correlated fluctuations across channels, and illumination changes can overshadow the subtle pulse signal. As a result, both techniques may extract components dominated by artifacts rather than cardiovascular activity [1], [3].

To address these limitations, de Haan and Jeanne introduced a chrominance-based method that leverages the distinct spectral behavior of pulse and motion artifacts across color channels [2]. By constructing signals that emphasize chromatic differences, their approach effectively suppresses in-phase noise while preserving pulse-related variations.

In this work, we evaluate the performance of chrominance-based and BSS-based methods under varying conditions, focusing on their robustness to motion and lighting variability. Our goal is to compare signal quality, accuracy, and practical applicability across approaches.

II. Materials and Methods

A. *Dataset description*

The dataset consists of raw RGB signal traces extracted from video recordings of 20 subjects. Each recording lasts approximately 600 seconds, with RGB data sampled at 115 Hz using a camera device. Rather than full video sequences, the dataset directly contains pre-extracted RGB time series.

Alongside the RGB data, a reference blood volume pulse (BVP) signal is provided for each subject. This signal was recorded with the same sampling rate (115 Hz).

B. *Data pre-processing*

Data preprocessing is a critical step in signal processing, especially in the context of remote photoplethysmography (rPPG), where raw data can be noisy and prone to errors. The primary objective of preprocessing is to clean the data, ensuring that the signal used for heart rate estimation is both accurate and reliable. In this section, we detail the preprocessing steps applied to the RGB signals (Red, Green, and Blue channels) extracted from the video data.

1) Jump Detection and Correction: RGB signals extracted from video recordings may exhibit abrupt, isolated amplitude changes that do not correspond to actual physiological variations. These artifacts can result from transient sensor instability, lighting fluctuations, or encoding errors, and may compromise the reliability of heart rate estimation.

To address this, a **correction procedure** is applied independently to each color channel to identify and eliminate such discontinuities. The method detects sudden amplitude shifts that exceed the normal variability of the signal. When a jump is detected, a compensatory adjustment is applied to the subsequent segment to restore continuity and preserve the overall trend of the signal.

2) Interpolation: After the jump correction step, each RGB signal is **interpolated** to ensure a uniform sampling rate of 115 Hz, as the original data may contain irregular time intervals between samples. *Spline* interpolation is used for this purpose, as it produces smooth, continuous curves that preserve the natural dynamics of the original signal.

Uniform resampling is essential in rPPG processing to ensure that frequency-domain analysis and filtering operations are accurate and reliable.

3) Low-pass Filtering: After interpolation, a low-pass filter is applied to each RGB channel to limit the frequency content of the signal. A fourth-order Butterworth filter with a cutoff frequency of 10 Hz is used.

This step is not strictly required for heart rate extraction. However, since the signals in the reference study were acquired at approximately 20 frames per second, the low-pass filtering is introduced here to match their effective bandwidth. This ensures greater consistency in spectral content and enables more reliable comparisons between methods.

C. *Chrominance-based Methods*

To estimate pulse rate from the preprocessed RGB signals, we implemented four chrominance-based rPPG algorithms. These methods aim to improve **robustness against motion artifacts and illumination variations** by leveraging the fact that pulse-induced color changes affect RGB channels differently from motion-induced fluctuations.

Chrominance-based approaches combine the color channels in specific ways to isolate the subtle, periodic chromatic variations caused by blood volume changes. Below we describe the four methods used and their implementation details.

1) RoverG: RoverG is a commonly used chrominance-based method for pulse rate estimation. It exploits the spectral characteristics of skin reflectance and hemoglobin absorption. Hemoglobin exhibits higher absorption in the green spectrum (around 540–580 nm), which makes the **green channel** particularly sensitive to blood volume changes. In contrast, the **red channel** is less affected by these fluctuations. By computing the ratio of normalized green to red intensity, the method amplifies pulse-induced chromatic changes while reducing common-mode noise, such as illumination variations:

$$S_{RoverG} = \frac{G_n}{R_n} - 1$$

where G_n and R_n represent the green and red channels normalized over time.

Normalization is performed using a sliding window of 1,6 second, where each sample is divided by the local mean intensity within that window:

$$C_n(i) = \frac{C(i)}{\mu_w(C(i))}$$

where $C(i)$ is the raw intensity of channel C at time i, and μ_w denotes the mean over the window. This normalization compensates for gradual illumination changes but does not eliminate fast variations within the window, which may still impact signal quality.

While RoverG is effective under controlled lighting conditions, it is sensitive to specular reflections—light directly reflected from the surface of the skin—which affect all RGB channels in a similar way and do not carry pulsatile information. This can degrade signal quality in dynamic environments with motion or illumination changes.

According to the Dichromatic Reflection Model (DRM), widely used in rPPG literature, the observed color signal can be decomposed into two components: a *diffuse reflection*, which is modulated by blood volume changes and thus contains physiological information; and a specular reflection, which is unmodulated and originates from the skin surface [10], [12].

Because RoverG relies on chrominance ratios between channels, it cannot effectively separate these two components. As a result, the presence of specular reflections can introduce non-physiological artifacts in the extracted signal, reducing the reliability of pulse rate estimation under uncontrolled conditions.

2) XoverY: The XoverY method addresses some of the limitations of RoverG by constructing two chrominance signals, X and Y, that emphasize color differences rather than relying solely on a direct ratio of channel intensities. Specifically, while RoverG is sensitive to uniform changes across RGB channels, XoverY improves robustness by isolating chromatic variations more closely linked to physiological activity. This formulation enhances robustness to motion artifacts and moderate lighting fluctuations.

The chrominance signals are defined as:

$$X = R - G, \quad Y = 0.5R + 0.5G - B$$

where R, G, and B represent the raw intensity signals of the red, green, and blue channels, respectively.

The signal $X = R - G$ captures the differential pulsatile behavior between the red and green channels, while $Y = 0.5R + 0.5G - B$ combines all three channels in a way that models global illumination changes and helps suppress motion-induced artifacts. The resulting ratio enhances pulse-related chromatic variations while minimizing non-physiological fluctuations.

The final rPPG signal is obtained from the normalized ratio:

$$S_{XoverY} = \frac{X_n}{Y_n} - 1,$$

where X_n and Y_n are the normalized versions of X and Y, computed using a sliding-window normalization identical to that used in the RoverG method. This step compensates for slow illumination drifts and enhances temporal stability.

Although XoverY is more robust than RoverG, it still has limitations. Under non-white or colored lighting conditions, the assumption that specular reflections affect all channels uniformly may no longer hold. Furthermore, rapid changes in illumination or skin reflectance can still introduce artifacts that degrade the accuracy of pulse rate estimation.

3) XoverY Fixed: The XoverY Fixed method extends the original XoverY formulation by incorporating a skin-tone standardization model. This approach compensates for distortions in the relative proportions of RGB channel intensities caused

by ambient lighting, particularly in non-white or artificial illumination conditions. By aligning the observed signals with a reference skin reflectance model under standard white light, the method enhances robustness to lighting variations.

The method applies fixed scaling coefficients derived from a standard white illumination model for human skin reflectance. The standardized channels are defined as:

$$R_s = 0.7682R_n, \quad G_s = 0.5121G_n, \quad B_s = 0.3841B_n,$$

where R_n, G_n, and B_n denote the red, green, and blue signals normalized using the same sliding-window approach as in the RoverG method. These normalized signals are then scaled using fixed coefficients to obtain standardized values, which are used to compute the corrected chrominance signals:

$$X_s = \frac{R_s - G_s}{0.7682 - 0.5121} = 3R_n - 2G_n,$$

$$Y_s = \frac{R_s + G_s - 2B_s}{0.7682 + 0.5121 - 0.7682} = 1.5R_s + G_s - 1.5B_s,$$

and the final pulse signal is extracted as:

$$S_{Fixed} = \frac{X_s}{Y_s} - 1.$$

By aligning the RGB channels to a canonical skin-tone model under white light, this method improves robustness to color distortions induced by ambient lighting. Consequently, XoverY Fixed yields more stable pulse estimation in challenging conditions, such as indoor environments or scenes illuminated by colored light sources.

4) $X_s min\alpha Y_s$: The $X_s min\alpha Y_s$ method builds upon the Fixed variant of the XoverY approach, which was introduced to improve robustness to motion and lighting changes by leveraging chrominance information. While the Fixed method compensates for lighting color using a standardized skin-tone model, it assumes that the chrominance signals X_s and Y_s contribute equally to the extracted pulse signal. In practice, their relative amplitudes may vary due to residual lighting effects or reflectance differences.

To address this, the $X_s min\alpha Y_s$ method introduces a dynamic scaling factor α, designed to balance the contribution of the two chrominance signals before pulse extraction. This facilitates a more stable and meaningful estimation of α by limiting the influence of broadband noise and slow drifts [2].

The correction factor is defined as:

$$\alpha = \frac{\sigma(X_f)}{\sigma(Y_f)},$$

where X_f and Y_f are the filtered versions of X_s and Y_s, and $\sigma(\cdot)$ denotes the standard deviation. The final rPPG signal is then obtained as:

$$S_{Xsmin\alpha Ys} = X_f - \alpha Y_f.$$

This formulation enables adaptive pulse extraction, where the relative influence of each chrominance component is adjusted dynamically to better reflect physiological variations and suppress non-informative fluctuations.

Rather than computing α over the entire signal, we implemented the method using an overlapping-add strategy, as proposed in the reference paper. Specifically, the signal is divided into overlapping segments using a 50% overlapping Hann window. Within each segment, a local value of α is estimated, and the corresponding pulse signal is computed as $X_f min\alpha Y_f$. The final output signal is then reconstructed by summing the contributions from each segment, weighted by the Hann window, following the procedure.

This strategy allows the method to adapt dynamically to local variations in motion or lighting. We tested multiple window lengths and evaluated performance by comparing the estimated pulse rate with the reference pulse rate derived from peak-to-peak intervals in the ground-truth BVP signal. Our results showed that shorter windows improved responsiveness and accuracy, while larger windows failed to capture rapid temporal changes. Based on these findings, we selected a window duration of approximately 1.6 seconds (32 samples), consistent with the original reference.

D. Blind Source Separation (BSS) Techniques

As part of our BSS-based rPPG pipeline, we explored two decomposition methods: Independent Component Analysis (ICA) and Principal Component Analysis (PCA). Both aim to isolate the pulsatile signal from RGB data, but rely on different statistical assumptions.

1) Independent Component Analysis (ICA): ICA decomposes multivariate signals into statistically independent components. In rPPG, the RGB channels are treated as linear mixtures of latent sources—ideally, one driven by cardiac-induced skin color variations, the others by noise or artifacts.

We adopt a preprocessing pipeline inspired by Poh et al. [7], starting with signal detrending via the Smoothness Priors Approach (cutoff frequency 0.89 Hz) to suppress baseline drift. Each channel is then normalized within overlapping Hann windows (50% overlap) to enforce zero mean and unit variance, improving ICA performance.

ICA is applied within each window, and the resulting components are analyzed spectrally. The one with the strongest peak within the heart rate band (40–240 BPM) is selected as the rPPG signal. To further reduce residual noise, a 5-point moving average is applied.

While ICA is effective under stable conditions, it faces limitations in real-world scenarios. First, its underlying assumption of statistical independence between sources often fails in the presence of motion or illumination changes, where sources become correlated [5]. Second, ICA relies on linear mixing, which does not always hold due to non-linear light–skin interactions [8]. Finally, selecting the pulsatile component based on spectral heuristics introduces subjectivity and may vary across subjects and lighting conditions.

2) Principal Component Analysis (PCA): Unlike ICA, PCA identifies uncorrelated (rather than independent) components by projecting the signal onto directions of maximum variance.

While computationally simpler, PCA assumes that the pulsatile signal is the dominant source of variance—an assumption that may not hold in real-world settings.

Our PCA implementation uses the same overlapping window structure. Prior to decomposition, a bandpass FIR filter is applied to the RGB signals to retain only the physiological band (40–240 BPM), removing unrelated low- and high-frequency components.

Spectral analysis is then performed on each principal component, and the one with the strongest peak in the target range is selected as the rPPG estimate.

PCA offers efficient pulse signal extraction under low-motion and well-lit conditions, but it can be misled when motion or illumination artifacts dominate variance [3], [6]. Additionally, the lack of enforced independence means the extracted component may combine multiple sources, reducing signal purity [3].

E. Processing of rPPG Signals and Spectral Analysis

Each extracted rPPG signal underwent a processing pipeline designed to enhance the pulsatile component and suppress noise and artifacts. This included bandpass filtering to isolate the physiological frequency range, temporal alignment with the reference BVP signal, and spectral analysis to identify dominant frequency components.

1) Filtering BVP and rPPG Signals: To ensure consistency in frequency content and avoid biases introduced by postprocessing, both the reference BVP signal and the rPPG signals were filtered using the same preprocessing steps.

A fourth-order Butterworth infinite impulse response (IIR) band-pass filter was applied to both the BVP and all rPPG signals (RoverG, XoverY Basic, XoverY Fixed, $X_s min \alpha Y_s$, ICA, and PCA). The filter was designed to retain only the physiological frequency range of interest (40–240 BPM), removing low-frequency drifts and high-frequency noise.

2) Synchronization of BVP and rPPG Signals: A key step in evaluating rPPG performance is ensuring temporal alignment between the extracted rPPG signals and the ground-truth BVP signal. Due to potential discrepancies in acquisition timing or signal latency, direct comparison without synchronization may lead to inaccurate assessments.

To address this, we performed a cross-correlation analysis between the BVP signal and the most reliable rPPG obtained using the $X_s min \alpha Y_s$ method. The lag corresponding to the maximum cross-correlation was identified and used to temporally align all rPPG signals to the BVP reference. This adjustment ensures that subsequent comparisons (e.g., spectral or time-domain) reflect true signal correspondence.

3) Spectral Analysis and Pulse Rate Estimation: To estimate the pulse rate from the extracted rPPG signals, we performed spectral analysis based on the Fast Fourier Transform (FFT). The objective was to identify the dominant frequency component corresponding to the subject's heart rate.

Each rPPG signal was divided into overlapping temporal segments of approximately 12 seconds. A Hann window was applied to each segment to reduce spectral leakage and improve frequency resolution. FFT was then computed for each window to obtain the power spectrum.

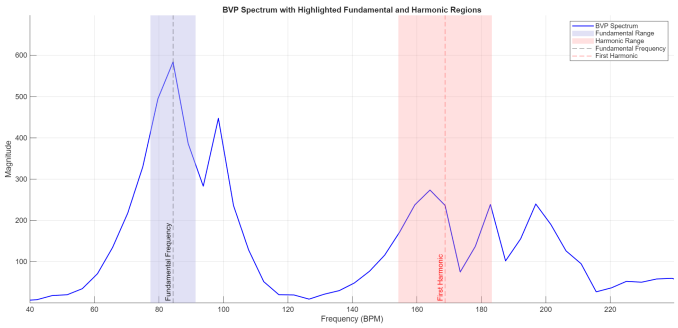

Fig. 1: BVP spectrum showing the fundamental frequency and harmonic range used for the template.

Within each spectrum, we searched for the dominant peak in the physiological frequency range of 40 to 240 BPM. The frequency corresponding to this peak was taken as the estimated pulse rate for that window. This approach allows for time-resolved tracking of pulse rate and enables comparison with the reference BVP signal.

4) Signal-to-Noise Ratio Evaluation: The signal-to-noise ratio (SNR) was used to quantitatively assess the quality of the extracted rPPG signals. The SNR was computed as:

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum (S_f \cdot T_f)^2}{\sum (S_f \cdot (1 - T_f))^2} \right) \tag{1}$$

where:

- S_f represents the spectrum of the signal,
- T_f is a template function calculated on the reference method that isolates the fundamental frequency and its harmonic.

The numerator accounts for the signal, while the denominator captures the noise outside the region of interest. This metric was computed for all rPPG methods and compared to the BVP signal.

5) Bland-Altman Analysis for Agreement Assessment: To assess the agreement between the pulse rates estimated from rPPG signals and the reference BVP signal, a Bland–Altman analysis was conducted. This statistical technique evaluates the consistency between two measurement methods by quantifying their mean difference (bias) and the range within which most differences lie (limits of agreement).

For each rPPG method, the difference between the estimated pulse rate and the corresponding reference BVP value was computed for all temporal segments. These differences were then plotted against the mean of the rPPG and BVP values for each segment. The bias (mean difference) and the limits of agreement (bias ± 1.96 times the standard deviation of the differences) were calculated and visualized. This analysis provides insight into the systematic error and variability associated with each method.

Furthermore, the percentage of data points falling within the Bland–Altman limits of agreement was computed for each

rPPG method. Under the assumptions of normally distributed differences and consistent measurement variability, approximately 95% of the differences are expected to fall within these limits. Therefore, a method is generally considered reliable if at least 95% of the data points lie within the limits of agreement [4].

III. Results

This section presents the performance of the rPPG methods using statistical, spectral, and signal-quality metrics. We evaluate global accuracy, agreement with the reference signal, and robustness across heart rate ranges and subjects.

A. Global Estimation Performances

Table I summarizes the performance of six rPPG extraction methods: *RoverG*, *XoverY*, *Fixed*, $X_s min\alpha Y_s$, *ICA*, and *PCA*. Four key evaluation metrics are reported: the Pearson correlation coefficient (r), the slope of the linear fit (B), the standard deviation of the residuals (σ), and the root-mean-square error (RMSE, ϵ).

Among all methods, *RoverG* and *PCA* exhibit the highest correlation with the ground truth signal ($r = 0.88$ and $r = 0.87$, respectively), along with the lowest RMSE values ($\epsilon = 4.75$ BPM and $\epsilon = 4.95$ BPM), indicating strong agreement with the reference. The chrominance-based method $X_s min\alpha Y_s$ performs better than its fixed-coefficient counterpart, with improved correlation and reduced error metrics.

Conversely, the *XoverY* method shows the weakest performance across all metrics ($r = 0.32$, $\epsilon = 20.83$ BPM), suggesting poor robustness to noise and varying lighting conditions. The *ICA*-based approach achieves only moderate correlation ($r = 0.36$) and exhibits higher variability in the estimated heart rate, likely due to sensitivity to motion artifacts and signal mixing limitations.

	RoverG	*XoverY*	Fixed	$X_s min\alpha Y_s$	*ICA*	*PCA*
r	0.88	0.32	0.65	0.71	0.36	0.87
B	0.77	0.13	0.48	0.52	0.23	0.76
σ	4.73	20.70	9.26	8.62	13.64	4.94
ϵ	4.75	20.83	9.41	8.67	13.92	4.95

TABLE I: Pearson correlation r, slope of linear fit B, standard deviation σ, and RMSE ϵ for each method.

B. SNR Evaluation

To assess the quality of the extracted pulse signals, the Signal-to-Noise Ratio (SNR) was computed for each rPPG method, according to what described in section II. Figure 2 visually compares the SNR values across the different approaches, while Table II reports the exact values in decibels (dB).

As shown in Table II, the $X_s min\alpha Y_s$ method achieved the highest SNR (22.07 dB), indicating superior signal clarity and noise suppression. The *PCA*-based method follows closely with an SNR of 21.87 dB. Conversely, the *XoverY* Basic and *ICA* methods exhibited the lowest SNR values (15.77 dB and 15.65 dB, respectively), suggesting a reduced ability to isolate the true heartbeat signal.

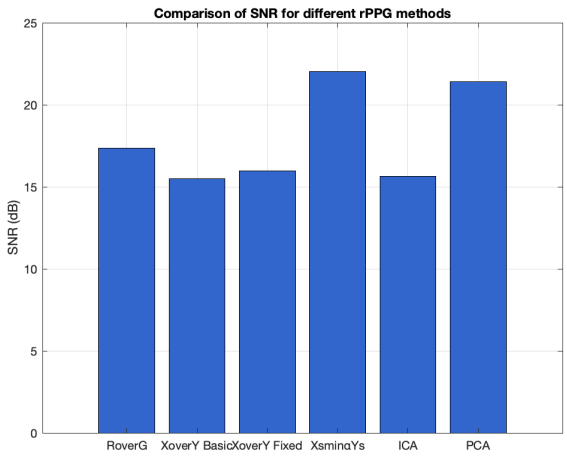

Fig. 2: mean SNR for different rPPG methods computed for 20 patients

TABLE II: Signal-to-Noise Ratio (SNR) for different rPPG methods

Method	**SNR (dB)**
RoverG	17.52
XoverY Basic	15.77
XoverY Fixed	16.03
XsminαYs	**22.07**
ICA	15.65
PCA	21.87

C. Agreement Analysis (Bland–Altman)

To assess the agreement between the estimated pulse rates and the reference BVP signal, a Bland–Altman analysis was conducted for each method. This technique allows the identification of systematic bias and the dispersion of measurement errors.

The analysis computes the **bias**, defined as the mean difference between the estimated and reference heart rates, and the **limits of agreement (LoA)**, calculated as:

$$\text{LoA} = \text{bias} \pm 1.96 \cdot \sigma$$

where σ is the standard deviation of the differences.

Table IV evaluates the agreement between pulse rate estimates from different methods and the reference BVP signal by reporting the bias, limits of agreement (LoA), and precision metrics.

The $X_s min\alpha Y_s$ method showed the best overall agreement, with a low bias of +0.87 BPM and tight LoA boundaries spanning from -16.03 to +17.78 BPM. Notably, it also achieved the highest percentage of pulse rate estimates within the LoA (98.38%) and a high precision of 87.62% within ± 5 BPM from the reference, demonstrating both accuracy and consistency.

Similarly, the *PCA* and *RoverG* methods displayed strong performance with biases near zero (+0.32 and -0.45 BPM, respectively) and narrow LoA intervals, alongside a high percentage of values contained within these limits (96.82% and 96.26%). Their precision within ± 5 BPM was also substantial, at 88.85% for *PCA* and 88.37% for *RoverG*.

In contrast, the *XoverY Basic* and *ICA* methods showed comparatively poorer agreement. Both exhibited larger biases (+2.32 BPM and -2.77 BPM) and considerably wider LoA ranges ([-38.26, +42.90] BPM and [-29.52, +23.97] BPM). These methods also had the lowest percentages of values falling within the LoA (93.67% and 93.15%) and relatively low precision within ± 5 BPM (75.04% and 75.18%), indicating less consistent estimates.

The Bland–Altman plot for the $X_s min \alpha Y_s$ method (Figure 4) confirms these findings, showing a bias close to zero and most points lying within the LoA. However, a trend toward greater deviations at higher mean heart rates is visible.

D. Subject-wise Performance

While global performance metrics provide an overall picture of each method's effectiveness, a more granular analysis reveals critical limitations related to heart rate dynamics and inter-subject variability.

A subject-wise analysis was performed to investigate the variability in estimation robustness across individual subjects. This approach helps to identify subjects who are consistently more challenging for the algorithms. Specifically, by examining the percentage of outliers beyond the LoA for each subject (excluding the reference BVP), it was possible to highlight a subset of subjects exhibiting systematically poorer performance across multiple methods. Most subjects, however, maintained low average outlier rates below 7%.

Subject 11 was the most problematic, with an average of 19.9% of estimates falling outside the LoA. Other notable cases include Subject 29 (8%) and Subject 20 (7.7%). These values may reflect consistent estimation difficulties, even when using otherwise robust methods.

To further understand whether a particular method was driving the overall mean error rates observed in some subjects, a detailed per-method analysis was performed. This breakdown allowed identification of methods that contributed most to the degraded performance in challenging cases.

The detailed per-method breakdown shows that for Subject 11 ICA was especially unreliable on this subject (37.0%), followed by $S_{X\text{overY}_b}$ (25.3%) and $S_{X\text{overY}_f}$ (21.6%). Even PCA and $S_{R/G}$ exceeded 14%, while the proposed $S_{X_s \min \alpha Y_s}$ method maintained a much lower error rate (4.7%).

Overall, the heatmap visualization in figure 3 underscores the heterogeneity of algorithm robustness. Some methods (e.g., $S_{X_s \min \alpha Y_s}$) display consistent performance across subjects, while others exhibit sporadic yet severe drops.

E. Heart Rate Regime Dependency

To better understand how estimation reliability varies with physiological dynamics, we analyzed the percentage of outlier pulse rate estimates for each method across three heart rate bands as shown in table III. Importantly, these HR bands were determined according to the ground-truth reference pulse rate, computed via RR peak detection.

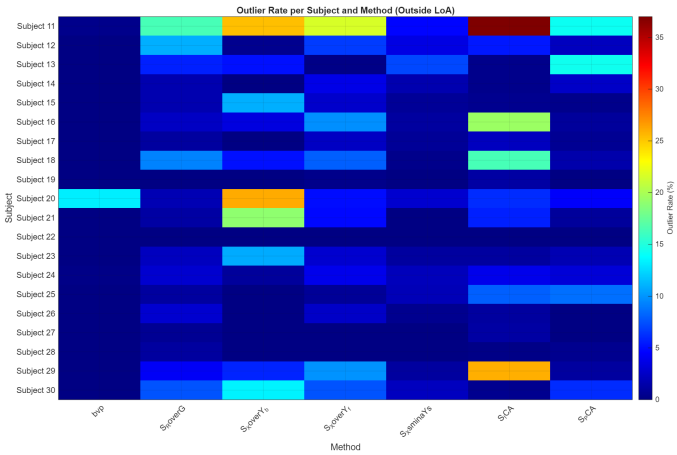

Fig. 3: Heatmap of outlier rates across subjects and methods. Each cell shows the percentage of pulse rate estimates falling outside the Bland–Altman limits of agreement (LoA) for a given subject (rows) and method (columns)

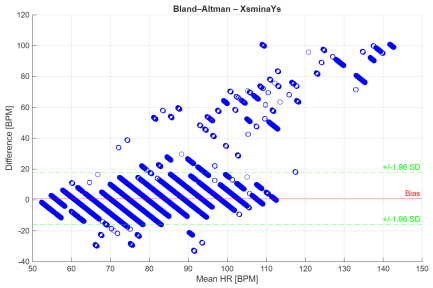

Fig. 4: Bland Altman plot for the XsminαYs method

TABLE III: Percentage of outliers per method across HR bands, as defined by the reference HR.

Method	Under 70 BPM	70–110 BPM	Over 110 BPM
RoverG	2.30%	3.97%	10.16%
XoverY$_b$	5.08%	6.38%	55.66%
XoverY$_f$	0.79%	5.22%	35.64%
X$_s$minαY$_s$	3.07%	1.36%	0.00%
ICA	0.70%	7.76%	61.17%
PCA	4.45%	2.96%	0.00%

Notably, many methods, including $XoverY_b$, $XoverY_f$, and especially ICA, exhibit severe degradation at actual higher heart rates. Conversely, methods such as PCA and the proposed $X_s min \alpha Y_s$ display robust behavior in the elevated heart rate range, with 0% of outlier estimates above 110 BPM. However, their relative performance deteriorates in the lower HR bands (e.g., 3.06% and 4.45% outliers under 70 BPM for $X_s min \alpha Y_s$ and PCA, respectively).

TABLE IV: Limits of Agreement (LoA) and Precision for Each Method

Method	Bias [BPM]	LoA Lower [BPM]	LoA Upper [BPM]	% Within LoA	% Within $\pm$5 BPM
RoverG	-0.45	-9.72	+8.82	96.26%	88.37%
XoverY_basic	+2.32	-38.26	+42.90	93.67%	75.04%
XoverY_fixed	-1.67	-19.83	+16.48	85.00%	83.29%
XsminaYs	+0.87	-16.03	+17.78	98.38%	87.62%
ICA	-2.77	-29.52	+23.97	93.15%	75.18%
PCA	+0.32	-9.35	+10.00	96.82%	88.85%

IV. Discussion

This section provides a critical interpretation of the results obtained and presented in Section III, focusing on the strengths and limitations of each rPPG method under test. We also analyze the trade-offs between accuracy, robustness, and signal quality, as well as implications for real-time applications.

A. Method Comparison and Trade-offs

The comparative analysis highlights trade-offs among the evaluated rPPG methods, each reflecting different design philosophies and operational assumptions.

The **RoverG** method demonstrates good global correlation with the reference signal, indicating that it can generally follow heart rate trends. However, its relatively low SNR suggests that the extracted PR remains affected by residual noise. This vulnerability becomes more evident under subject-specific variability, such as motion, lighting changes, or skin-tone differences, where the method shows increased outlier rates. Since RoverG relies on a fixed RGB ratio without any form of adaptive compensation, its robustness is inherently limited. While it may perform adequately in stable and controlled conditions, its effectiveness diminishes in more dynamic or unconstrained scenarios.

The **XoverY Basic** method performs the weakest among all tested approaches, likely due to its reliance on fixed chromatic channel ratios without normalization on the channels. The resulting estimates are highly sensitive to inter-subject variability, motion, and illumination changes—leading to extremely wide limits of agreement (LoA) and the highest RMSE observed. These results highlight the limitations of using static linear combinations of color channels in real-world conditions, where the spectral and spatial characteristics of the signal are inherently dynamic.

The **Fixed chrominance ratio** variant (*XoverY Fixed*) shows partial improvement over the basic formulation, confirming that chrominance-based representations can enhance robustness compared to naive color division. However, it still shows limited adaptability. This is particularly evident at elevated heart rates, where motion-induced spectral components become more prominent, leading to increased error rates and reduced reliability.

In contrast, the adaptive $X_s \min \alpha Y_s$ method demonstrates consistently strong performance across all evaluation criteria. Its high SNR, narrow LoA, and stable subject-wise reliability indicate that the adaptive chrominance normalization,implemented via the time-varying coefficient α, successfully mitigates the effects of motion and illumination variability.

The **ICA**-based method exhibits only modest performance, with low correlation and poor signal quality. These shortcomings are consistent with the known limitations of blind source separation techniques under non-ideal conditions. Specifically, ICA relies on assumptions of statistical independence and linear, stationary mixing—conditions rarely satisfied in facial video, where physiological signals are often entangled with motion, lighting, and compression artifacts.

By contrast, the **PCA**-based method outperforms ICA and closely matches $X_s \min \alpha Y_s$ in terms of correlation and RMSE. This suggests that, in this dataset, the cardiac signal often aligns with a dominant axis of variance in the RGB space. Unlike ICA, PCA does not require independence, and can effectively capture correlated oscillations across channels. Nonetheless, its performance is not uniformly stable across all subjects, likely due to inter-individual differences in signal mixing or noise characteristics.

Overall, the results confirm that methods incorporating temporal adaptivity or leveraging statistical structure—such as $X_s \min \alpha Y_s$ and PCA—offer superior robustness in the face of real-world variability.

B. Impact of Subject Variability and Heart Rate regimes

The performance of rPPG methods shows non-negligible variation both across individuals and HR regimes. These discrepancies reveal limitations not fully captured by global metrics. Hence, we conducted an in-depth analysis.

a) Inter-subject variability.: As shown in Figure 3, most subjects maintain a low percentage of outliers, but several individuals, such as Subject 11, Subject 20, and Subject 29—exhibit systematically degraded performance across multiple methods, as presented in section III.

Such failures may originate from a range of uncontrolled factors: skin tone, facial geometry, facial motion (e.g., blinking, speech), ambient lighting changes, or acquisition artifacts such as sensor defocus.

In some instances, even the reference BVP signal shows signs of degradation. In particular, Subject 20 exhibited 13.2% of LoA violations in the BVP-derived signal itself—raising concerns about the reliability of ground truth. Figure 5 compares two HR estimations from BVP: one based on RR intervals (red) and one from spectral peak detection (green). While the RR-based curve remains physiologically plausible, the spectral method displays erratic spikes above 200 BPM.

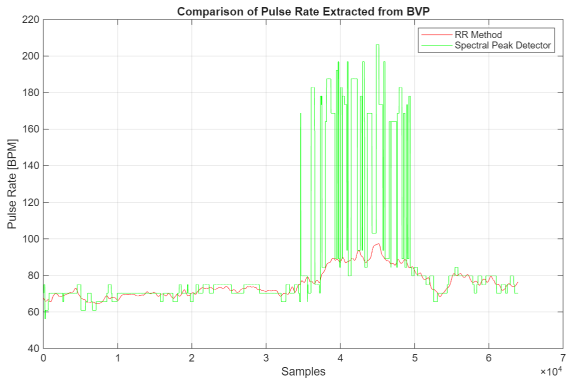

Fig. 5: Comparison between HR estimates derived from BVP signal: RR interval-based (red) versus spectral peak-based (green).

These findings suggest that the poor performance observed in some cases may be attributed not only to the estimation methods but also to the quality of the signal itself. Notably, when peak detection algorithms applied to the contact-based BVP signal yield physiologically implausible heart rate estimates, this likely indicates underlying signal degradation. Such degradation can stem from challenging conditions like significant motion artifacts, sensor displacement, or suboptimal skin contact.

In these scenarios, the reference BVP signal already suffers from compromised quality, making it inherently difficult to obtain accurate heart rate measurements. Consequently, estimating heart rate from rPPG signals — which are generally more susceptible to noise and motion artifacts — becomes even more challenging. This underscores the importance of carefully assessing the reliability of the reference signal before interpreting the performance of rPPG estimation methods.

b) Heart rate regime dependency: The methods also show clear performance differences across heart rate (HR) bands. As shown in Table III, most techniques exhibit notable degradation at when the actual heart rate is high (>110 BPM), particularly ICA (61.2% outliers) and *XoverY Basic* (55.7%). This effect is consistent with previous observations in the literature [2], [10], where high HR regimes were found to be especially vulnerable to motion-induced spectral interference.

Since the rPPG signal is extracted from RGB video, even subtle head movements or facial expressions can introduce periodic components within the same frequency range as the heart rate, especially during activities like speaking, nodding, or shifting posture. Moreover, the frequency resolution of the Fourier spectrum becomes limited when using short time windows, which are necessary for real-time tracking. As the heart rate increases, the spacing between harmonics widens, but the reduced resolution may prevent the clear separation of closely spaced peaks caused by motion and pulse signals. This issues are particularly problematic for methods relying solely on a single dominant spectral component, without leveraging temporal continuity or spatial constraints [12].

Conversely, both PCA and $X_s \min \alpha Y_s$ exhibit robust behavior in the high-HR regime, with 0% of estimates classified as outliers above 110 BPM (Table III). This result suggests that these methods are particularly effective at suppressing motion-related spectral components that commonly interfere in this range. In the case of PCA, this robustness likely stems from its ability to emphasize consistent signal components across color channels while attenuating uncorrelated motion artifacts. Similarly, $X_s \min \alpha Y_s$ applies a dynamic chrominance normalization that can suppress illumination- or motion-induced imbalances, which tend to dominate at higher frequencies.

However, an opposite trend is observed at lower heart rates (<70 BPM), where both methods show non-negligible error rates (3.07% for $X_s \min \alpha Y_s$, 4.45% for PCA). The possible causes of this performance degradation are explored in detail in the following subsection.

Overall, the findings support the view that both inter-subject and frequency-dependent variability must be explicitly addressed. Average-case performance metrics can mask critical failure cases, especially under dynamic or adverse conditions.

C. Spectral Behavior and Harmonics

A detailed spectral analysis was performed to investigate the origin of the largest HR estimation errors observed in the $X_s \min \alpha Y_s$ method, particularly in conditions where the true pulse frequency is low. As noted in earlier sections, while this method generally maintains high robustness, it occasionally produces extreme outliers.

Closer inspection of these anomalous cases revealed a recurring pattern: the algorithm often misidentifies harmonic components, typically the second harmonic, as the fundamental frequency. In these instances, the spectral component at the second harmonic exhibits higher amplitude than the true fundamental, misleading the peak-selection mechanism.

This issue is particularly problematic when the harmonic frequency coincides with a physiologically plausible heart rate range (e.g., 120–200 BPM), causing the error to appear realistic even though it reflects a misidentification. Such conditions frequently arise during motion, motion-induced blurring, or uneven illumination—factors that can amplify or distort specific spectral components in the rPPG signal.

This type of harmonic misclassification has been noted in the rPPG literature, particularly under low-SNR or motion-contaminated conditions [9], [11]. In such cases, nonlinear effects, such as those caused by illumination variations or skin reflectance, may lead to stronger second harmonics that dominate the spectrum, misleading peak selection algorithms based on maximum amplitude.

Figure 6 shows a representative example of this phenomenon for Subject 14, in the time window corresponding to the maximum estimation error. Here, the peak around 135 BPM (in red) was incorrectly selected as the fundamental, whereas the true HR likely corresponds to the smaller but correctly located peak at 65 BPM (shown in black).

These findings underscore a fundamental limitation of peak-based spectral selection methods for $X_s min \alpha Y_s$: when

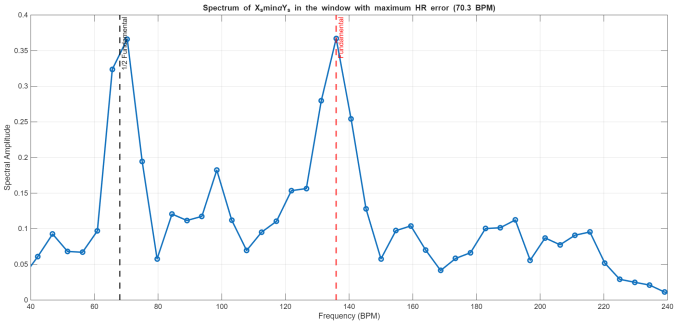

Fig. 6: Spectral representation of the $X_s \min \alpha Y_s$ signal in the time window corresponding to the maximum HR estimation error (70.3 BPM) for Subject 14. The detected (erroneous) fundamental frequency is highlighted in red, while the peak at approximately half that value, likely the true HR, is shown in black.

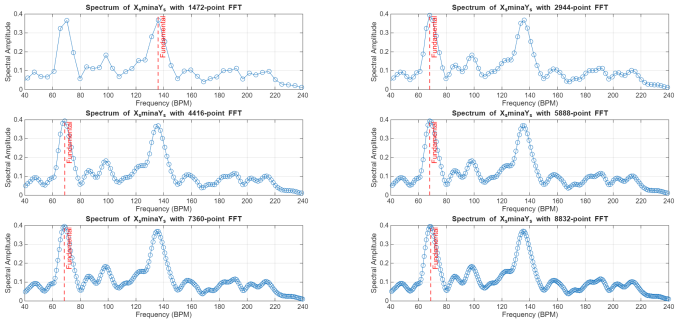

Fig. 7: Effect of FFT resolution on the power spectrum of the $X_s \min \alpha Y_s$ signal. Increased zero-padding sharpens spectral peaks and improves resolution of closely spaced components.

harmonics dominate the spectrum, they can easily lead to substantial misclassifications. This motivates the need for more sophisticated frequency analysis approaches.

To further understand the spectral behavior of the $X_s \min \alpha Y_s$ method and its vulnerability to harmonic misclassification we performed a systematic analysis of two key frequency-domain parameters: **FFT resolution** and **spectral windowing**. These components may directly influence the shape and interpretability of the power spectrum and can either mitigate peak selection errors in noisy or motion-contaminated conditions.

a) FFT Resolution: Increasing the number of FFT points via zero-padding enhances frequency resolution, allowing for finer discrimination between closely spaced components. Figure 7 shows the effect of increasing the FFT size (from 1472 to 8832 points) on the spectrum of a representative signal segment where the first harmonic was confused with the fundamental frequency. As resolution increases, peaks become clearer and side lobes more visible.

Despite experimenting with various zero-padding ratios to enhance spectral resolution, our results indicate that while zero-padding can improve the visual clarity of certain windows (see Figure 7), it does not lead to a significant overall performance gain. This confirms that the core issue lies in the intrinsic quality of the spectrum rather than its resolution. In low-SNR conditions, the true fundamental frequency often remains suppressed or obscured by noise and artifacts. Furthermore, higher FFT resolution comes at the cost of increased computational load and reduced temporal responsiveness, which may limit the feasibility of real-time applications.

b) Spectral Windowing.: The choice of spectral windowing function significantly impacts the frequency representation of rPPG signals. Figure 8 illustrates the power spectra obtained from the $X_s \min \alpha Y_s$ signal segment using six standard window types.

Among the tested functions, **Hann** and **Hamming** windows offer a good compromise: their moderate side-lobe attenuation and narrow main lobes preserve peak clarity while limiting spectral leakage. These characteristics make them well suited for rPPG scenarios where the fundamental frequency is present but may be surrounded by physiological harmonics or motion-induced noise.

Chebyshev and **Blackman** windows exhibit strong side-lobe suppression, which can help reduce contamination from nearby spectral components. In cluttered spectra, this can facilitate more robust peak isolation, especially when the fundamental is not the strongest component. However, the higher attenuation can sometimes mask low-amplitude harmonics or narrow peaks if the signal is weak.

Conversely, **Bartlett** and **Kaiser** windows yield broader spectral peaks with greater energy spreading. This can lead to peak blurring and makes it more difficult to distinguish the true heart rate from adjacent frequencies.

These observations underscore the importance of selecting an appropriate windowing function in rPPG systems, especially for real-time or short-window spectral estimation. The balance between resolution, leakage, and side-lobe control can significantly influence the accuracy and robustness of pulse rate detection. In our experiments, multiple window types were tested, and the **Hann** window provided the best trade-off between computational cost and spectral estimation quality.

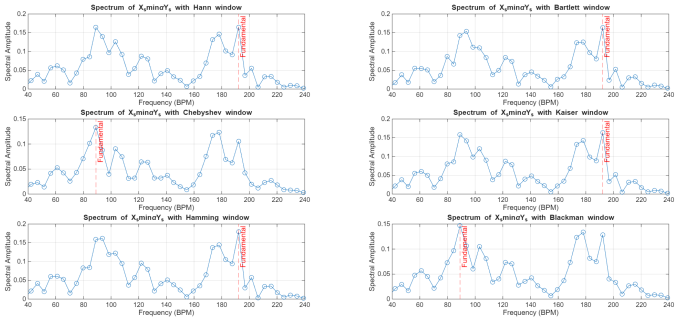

Fig. 8: Impact of different window functions on the spectral shape of the $X_s \min \alpha Y_s$ signal. Windowing affects the sharpness and visibility of harmonic components, directly influencing peak detection.

In summary, despite these differences, neither higher resolution via zero padding nor advanced windowing fully mitigates

the issue of harmonic dominance. This highlights a broader limitation of peak-based spectral estimation: when harmonics are amplified by motion, lighting variation, or nonlinear reflectance effects, they may dominate the spectrum and lead to large errors.

D. Role and Optimization of the Normalization Window

A key component of the $X_s min\alpha Y_s$ method is the normalization window used to compute the adaptive coefficient α, which dynamically balances the chrominance channels to suppress motion-induced distortions.

To assess the impact of this window length, a parametric evaluation was performed by varying the normalization window size and computing the Mean Absolute Error (MAE) between the estimated pulse rate and the ground-truth reference. The results in Figure 9, show that the MAE remains relatively stable and and small for shorter windows, indicating effective adaptation to local RGB variations without losing the pulsatile signal integrity.

However, beyond few seconds, the MAE begins to increase gradually, as the window length increase. This trend suggests that longer windows lead to excessive temporal smoothing, causing the coefficient α to become less responsive to fast temporal changes induced by motion, facial expressions, or lighting fluctuations. Consequently, the method's ability to accurately correct these disturbances diminishes, resulting in increased error.

This clear sensitivity of the MAE to the normalization window length highlights that the window size is not a trivial implementation detail but a critical design parameter, particularly in dynamic acquisition settings with significant motion or non-uniform illumination.

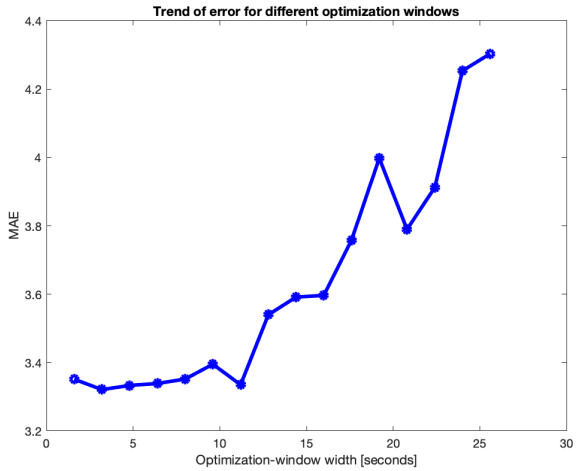

Fig. 9: Mean Absolute Error as a function of normalization window length for the $X_s \min \alpha Y_s$ method.

E. Comparison with the Reference Study

A comparative look at the original work by de Haan and Jeanne [2] provides important context for interpreting our findings, despite notable differences in experimental conditions and evaluation protocols.

In [2], the chrominance-based method $X_s \min \alpha Y_s$ was evaluated in controlled scenarios involving predefined physical activities such as stationary cycling and stepping. Under these settings, the method achieved over 98% accuracy (within ± 3 BPM) on the bike sequence using short temporal windows (1.6 s), significantly outperforming blind source separation methods like ICA and PCA, which achieved only 54% and 11% accuracy, respectively, even with longer analysis windows (25 s).

By contrast, our dataset lacks detailed annotations about subject activity or environmental conditions, which prevents a direct comparison under controlled motion. However, per-subject analysis of the reference BVP signals reveals heterogeneous dynamics across the recordings: some subjects appear to be at rest, while others exhibit abrupt heart rate changes. These observations suggest that our recordings encompass a broader range of unconstrained conditions, including naturalistic motion and noise.

Despite this variability, the $X_s min\alpha Y_s$ method consistently demonstrated strong performance in our evaluation, with results qualitatively aligning with those reported in [2]. This further underscores the robustness of chrominance-based approaches, particularly when short normalization windows (1–5 s) are employed, as confirmed by our parameter optimization analysis (see Figure 9).

Interestingly, unlike the findings in [2], where PCA showed marked degradation in high-motion sequences, our results indicate that PCA performs comparably to $X_s \min \alpha Y_s$ across several key metrics. This discrepancy may reflect differences in the nature or intensity of motion across datasets. In our recordings, the independent pulsatile component extracted by PCA may remain aligned with the true cardiac signal if motion artifacts are weak, unsynchronized, or orthogonal to the primary sources of variance. While PCA lacks explicit motion compensation mechanisms, its empirical success in our setting suggests that, under certain conditions, linear decomposition may still extract clean rPPG signals. This stands in contrast to some motion scenarios, where structured activities (e.g., stepping) likely introduce more coherent and disruptive artifacts that interfere with unsupervised separation.

In conclusion, although our study cannot replicate the tightly controlled environment of [2], our results reinforce its core insights: chrominance-based approaches—particularly those employing adaptive normalization—offer greater resilience to environmental variability. At the same time, the relatively strong performance of PCA under moderate motion highlights the importance of dataset characteristics in determining method effectiveness.

V. Conclusions and Future Work

This study conducted a comprehensive evaluation of six rPPG methods under unconstrained conditions, focusing on robustness, signal quality, and sensitivity to subject- and frequency-specific variability.

Our results show that static or blind approaches, such as *XoverY Basic* and ICA, are particularly vulnerable to motion artifacts, illumination changes, and individual differences. These methods exhibit high error rates, especially in high heart rate regimes, where spectral overlap with motion components is most pronounced.

In contrast, techniques that integrate temporal adaptivity or exploit statistical structure, such as $X_s \min \alpha Y_s$ and PCA, demonstrate significantly higher robustness. The $X_s \min \alpha Y_s$ method, in particular, benefits from dynamic chrominance normalization and achieves state-of-the-art performance across most metrics. However, detailed spectral analysis reveals that even this method can fail under certain conditions, notably when harmonics dominate the spectrum or when the normalization window is not optimally tuned. PCA, while lacking explicit compensation mechanisms, surprisingly shows strong empirical performance, likely due to favorable variance alignment in the dataset and minimal motion coupling.

The analysis also highlights critical challenges in rPPG evaluation: inter-subject variability and the reliability of reference BVP signals can strongly influence performance interpretation. In few cases, reference signals showed artifacts or inconsistencies.

The results of this study suggest several possible directions for future work, aimed at improving the reliability of pulse rate estimation and gaining deeper insight into the limitations of the tested methods.

1) **Post-processing techniques for correcting unreliable estimations.** In some cases, the estimated pulse rate deviates significantly from the reference signal, often due to the selection of a harmonic instead of the fundamental frequency or the presence of motion artifacts. A useful extension would be to introduce a post-processing stage capable of automatically detecting such anomalies and correcting the final estimation, for instance using temporal consistency checks or simple heart rate tracking models.
2) **Dataset expansion and acquisition condition analysis.** Increasing the number and diversity of subjects would improve the robustness of the results and allow better generalization. Moreover, having access to detailed acquisition information, such as lighting conditions, head movement, or skin tone, would make it possible to correlate performance drops with specific sources of variability and better understand the causes behind certain estimation failures.
3) **Comparison with more recent rPPG methods.** While this work focused on classical, interpretable chrominance-based techniques, it would be valuable to compare these approaches with more recent methods, including deep learning-based models.

REFERENCES

[1] Anantyananda Dasari, Sakthi Kumar Arul Prakash, László A Jeni, and Conrad S Tucker. Evaluation of biases in remote photoplethysmography methods. *NPJ digital medicine*, 4(1):91, 2021.

[2] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE transactions on biomedical engineering*, 60(10):2878–2886, 2013.

[3] Uday Debnath and Sungho Kim. A comprehensive review of heart rate measurement using remote photoplethysmography and deep learning. *BioMedical Engineering OnLine*, 24(1):1–34, 2025.

[4] Davide Giavarina. Understanding bland altman analysis. *Biochemia Medica*, 25(2):141–151, 2015. LoA contain approximately 95

[5] Byung Hoon Kim and Seung-Chan Yoo. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Transactions on Biomedical Engineering*, 53(3):566–568, 2006.

[6] Nhi Nguyen, Le Nguyen, Honghan Li, Miguel Bordallo, and Constantino Álvarez Casado. Evaluation of video-based rppg in challenging environments: Artifact mitigation and network resilience. *arXiv*, 2024.

[7] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.

[8] David Pollreisz and Nima TaheriNejad. Detection and removal of motion artifacts in ppg signals. *Mobile Networks and Applications*, 27(2):728–738, 2022.

[9] Dangdang Shao, Chenbin Liu, and Francis Tsow. Noncontact physiological measurement using a camera: a technical review and future directions. *ACS sensors*, 6(2):321–334, 2020.

[10] Willem Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445, 2008.

[11] Wenjin Wang, Geetha Balakrishnan, Fredo Durand, and Polina Golland. A novel algorithm for remote photoplethysmography using spatial subspace rotation. *IEEE Transactions on Biomedical Engineering*, 65(9):1995–2004, 2018.

[12] Wentao Wang, Geetha Balakrishnan, Hao Xu, Yuan Ma, and Zhi Zhang. Algorithmic principles of remote ppg. In *IEEE Transactions on Biomedical Engineering*, volume 64, pages 1479–1491, 2017.