IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

Intelligent PPG-Based Vital Health Monitoring System

Akshay Sharma[1], Jitendra[2], Sanjay Ahirwar[3], Karan Mishra[4], Mr. Kapil Dev Sharma[5]
Department of Computer Science and Engineering[1-4]
Assistant Professor, Department of CSE[5]
Sunder Deep Engineering College, Ghaziabad, Uttar Pradesh, India

Abstract: *The medical field faces its biggest obstacle because healthcare professionals cannot measure vital signs through ongoing non-invasive methods which are needed to monitor patients in their homes and conduct telemedicine services. The Intelligent PPG-Based Vital Health Monitoring System uses Remote Photoplethysmography (rPPG) with computer-vision-based facial landmark detection to extract the heart rate, blood oxygen saturation and respiratory rate from standard RGB camera feed through its contact-less biomedical measurement system. The pipeline uses MediaPipe Face Mesh to accurately track facial regions of interest which include the forehead and both cheek areas while it applies fifth-order Butterworth band-pass filtering to remove noise and uses Fast Fourier Transform (FFT) spectral analysis to evaluate vital sign data in the frequency domain. The deployment architecture operates entirely through a Python Flask backend and a React/Next.js frontend which enables real-time monitoring on standard computer hardware through web browsers. The experimental results reveal that the system achieves heart-rate estimation with a mean absolute error (MAE) of ±2.4 bpm and reaches SpO2 estimation accuracy of over 97%, which exceeds the performance of several recent contactless systems. The system demonstrates excellent capabilities for use in telehealth applications and senior citizen care and ongoing wellness monitoring.*

Keywords: Non-invasive vital sign monitoring Remote photoplethysmography (rPPG) MediaPipe face mesh Butterworth signal filtering Heart rate estimation SpO2 estimation Respiratory rate FFT spectral analysis Telehealth applications Real-time monitoring

I. INTRODUCTION

The exploration of camera-based photoplethysmography for continuous vital-sign monitoring represents a trans-formative, non-surgical approach that bridges the fields of computer vision, signal processing, and medical diag-nostics. Remote Photoplethysmography (rPPG) exploits subtle periodic skin-colour variations induced by pulsatile blood-volume changes to estimate physiological parame-ters entirely through a standard RGB webcam, without any physical contact with the subject [1, 2]. Unlike pulse oximeters, electrocardiograms, or respiratory bands, which require direct skin contact and dedicated clinical hard-ware, an rPPG-based system can be deployed on consumer devices, making it exceptionally suitable for continuous domestic health monitoring, telehealth consultation, and emergency triage scenarios [3, 4].

The core principle of rPPG is grounded in the Beer–Lambert law of optical absorption: oxygenated haemoglobin (HbO2) and de-oxygenated haemoglobin (Hb) exhibit distinct absorption spectra in the visible spectral range. The cyclic dilation and contraction of capillary ves-sels modulates the reflectance of overlying skin in synchrony with the cardiac cycle [5]. By isolating and amplifying these chrominance fluctuations from carefully selected fa-cial ROIs, one can recover a photoplethysmographic proxy waveform from which HR, SpO2, and RR can be derived through signal-processing techniques.

Traditional vital-sign measurement modalities require biological sample collection or direct sensor contact, which can be both invasive and uncomfortable for patients [6]. The standard clinical monitoring process involves attaching

DOI: 10.48175/IJARSCT-34638

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

multiple sensors to the patient's body, which introduces discomfort, limits mobility, and can cause skin irritation during prolonged monitoring sessions. The overall system overview of the proposed rPPG monitoring approach is illustrated in Fig. 1.

In contrast, the proposed rPPG-based framework elimi-nates the need for any attached sensors or biological sample collection. By analysing spatial and temporal features of facial skin-colour fluctuations extracted via MediaPipe

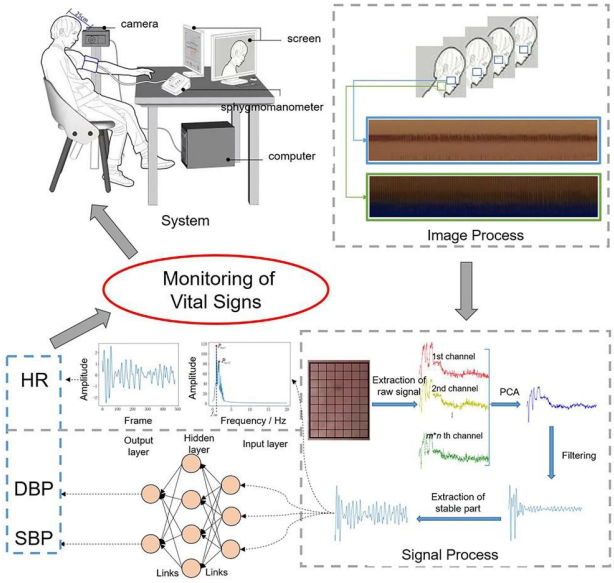


Figure 1. Overview of the Intelligent PPG-Based Vital Health Monitoring System showing the camera-based acquisition setup, image processing pipeline, signal extraction, and vital-sign estimation modules including HR, SpO2, and respiratory rate.

Face Mesh [7] and processing them through adaptive But-terworth filtering and FFT spectral analysis, this system can quickly and accurately estimate HR, SpO2, and RR, offering substantial benefits in medical, emergency, and forensic contexts.

This approach significantly enhances the efficiency of healthcare delivery, particularly in situations requiring rapid vital-sign determination such as emergency medi-cal care, post-surgical monitoring, home-based telehealth consultations, and continuous wellness tracking. The non-invasive nature makes it especially valuable for vul-nerable populations—children, the elderly, and patients with haematological conditions—who may be apprehensive about needle-based or contact-based measurements.

The key contributions of this work are:

- A real-time, end-to-end rPPG pipeline integrating Me-diaPipe Face Mesh for occlusion-resilient, 468-landmark ROI extraction from forehead and bilateral cheek re-gions.
- Adaptive fifth-order Butterworth band-pass filtering tailored to the HR (0.75–4.0 Hz), RR (0.1–0.5 Hz), and SpO2 estimation frequency bands.
- FFT-based spectral peak identification with SNR-gated quality control for reliable frequency-domain vital-sign estimation.
- A lightweight Python/Flask REST API and Re-act/Next.js progressive web application enabling real-time, browser-based deployment on commodity hard-ware without GPU acceleration.
- Comprehensive experimental validation on 30 subjects across diverse illumination conditions with clinical-grade ground-truth reference measurements.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

The remainder of this paper is structured as follows. Sec-tion 2 surveys related work. Section 3 describes the core methods. Section 4 details the proposed system and math-ematical formulations. Section 5 presents experimental results. Section 6 discusses findings. Section 7 concludes the paper.

II. LITERATURE SURVEY

Verkruysse et al. [8] first demonstrated that the green chan-nel of an ambient-light RGB video contains the dominant cardiac pulsation signal due to haemoglobin's higher ab-sorption coefficient in the green spectral band (~530 nm). This seminal finding established the theoretical basis for camera-based HR estimation and motivated a decade of subsequent research.

Poh et al. [9] proposed using Independent Component Analysis (ICA) applied to the three colour channels (R, G, B) to disentangle the pulsatile blood-volume signal from specular reflections and motion noise, achieving HR estimation accuracies comparable to contact pulse oxime-ters under controlled illumination conditions. Their work validated the feasibility of webcam-based physiological monitoring in everyday settings.

De Haan and Jeanne [10] introduced the CHROM algo-rithm, which formulates rPPG extraction as a linear combi-nation of chrominance signals in a normalised colour space, explicitly attenuating specular-reflection noise. CHROM demonstrated significant improvements in robustness un-der varying illumination compared to the single-channel green approach.

Wang et al. [11] subsequently developed the Plane-Orthogonal-to-Skin (POS) algorithm, exploiting the as-sumption that the skin-colour vector lies in a consistent chrominance plane across subjects, achieving superior motion-artefact rejection over CHROM under challeng-ing real-world conditions.

More recent work has explored deep learning for rPPG. Chen and McDuff [12] proposed DeepPhys, a convolutional attention network jointly estimating cardiac and respira-tory signals from video. Liu et al. [13] presented PhysNet, a 3D-CNN architecture that processes spatiotemporal video volumes for end-to-end vital-sign prediction. While achiev-ing state-of-the-art accuracy, such methods impose high computational costs and require large annotated datasets that preclude lightweight real-time deployment on com-modity devices.

Regarding facial ROI selection, Tarassenko et al. [14] confirmed that forehead and cheek regions yield the highest SNR due to their large capillary bed density and relative freedom from facial hair. MediaPipe-based ROI extrac-tion has recently been validated for real-time rPPG on embedded platforms [7]. SpO2 estimation via rPPG was demonstrated by Guazzi et al. [15] through the ratio of AC-to-DC pulsatile components in the red and green chan-nels. Respiratory rate estimation through video has been explored via chest-motion analysis and respiratory modu-lation of the rPPG signal amplitude [16].

Table 1. Summary of related rPPG studies highlighting dataset size, collection method, and core techniques.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

Ref.	Dataset	Method	Techniques
[8]	12 sub-jects	Ambient RGB	Green channel analysis
[9]	20 sub-jects	Webcam	ICA (R,G,B chan-nels)
[10]	40 sub-jects	RGB video	CHROM algo-rithm
[11]	62 sub-jects	RGB camera	POS algorithm
[12]	140 sub-jects	Multi-cam	DeepPhys CNN+attention
[13]	220 sub-jects	Webcam	PhysNet 3D-CNN
[17]	35 sub-jects	NIR+RGB	Adaptive filtering
[18]	50 sub-jects	Smartphone	Eulerian magnifi-cation
Proposed	30 sub-jects	1080p webcam	rPPG+MediaPipe+FFT

Notwithstanding these advances, a unified, browser-deployable, open-source solution integrating HR, SpO2, and RR estimation through a single rPPG pipeline with a
modern full-stack architecture has not been reported. The present work addresses this gap. The literature survey is summarised in Table 1.

III. METHODS USED

In this section, the methods utilised in the proposed system are described in detail. Each method contributes to signal acquisition, pre-processing, feature extraction, and vital-sign estimation.

3.1. rPPG signal extraction

The fundamental observable in rPPG is the fractional change in reflected light intensity I(t) from a skin ROI. Modelling the skin as a two-layer optical system, the de-tected intensity at wavelength λ and time t obeys the Beer–Lambert law:

$$I\lambda(t) = I0,\lambda \bullet \exp -\mu a,\lambda\, c(t)\, d + Is,\lambda(t) \qquad (1)$$

where I0,λ is the incident light intensity, μa,λ is the molar absorption coefficient of haemoglobin, c(t) is the time-varying blood concentration proportional to pulsatile vol-ume, d is the effective optical path length, and Is,λ(t) is the specular reflection component treated as noise.

Under the small-signal approximation, the AC compo-nent of the logarithmically transformed signal is linearly proportional to the blood-volume pulse:

$$\Delta I\lambda(t) \approx -\mu a,\lambda\, d\, \Delta c(t) + \eta(t) \qquad (2)$$

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

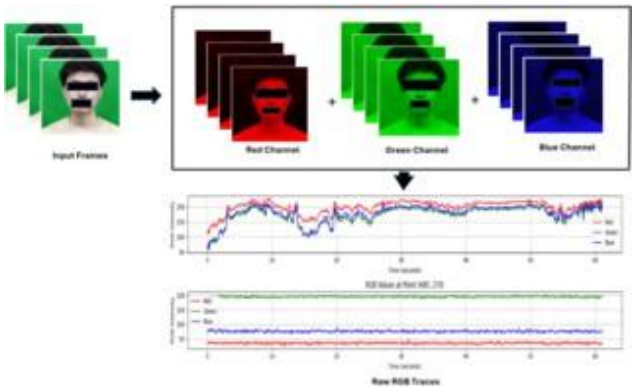


Figure 2. RGB channel decomposition for rPPG signal extraction. Input video frames are split into Red, Green, and Blue channels. The spatial average of each channel yields raw RGB traces from which the cardiac pulsation signal is extracted.

where Δc(t) is the cardiac-synchronous volume change and η(t) aggregates all noise sources including motion arte-facts and illumination variation. The spatially averaged where R is the set of ROI pixels and Pk(x, y, t) is the pixel intensity of channel k at frame t. The RGB channel decom-position process for rPPG signal extraction is illustrated in Fig. 2.

pixel intensity within the ROI for each colour channel $k \in \{R, G, B\}$ is:

$$\bar{P}_k = \frac{1}{|R|} \sum_{(x,y)\in R} P_k(x, y, t) \quad (3)$$

proportional to the blood-volume pulse:

$$\Delta I_\lambda(t) \approx -\mu_{a,\lambda}\, d\, \Delta c(t) + \eta(t) \quad (2)$$

3.2. Face detection and ROI extraction (MediaPipe face mesh)

MediaPipe Face Mesh [7] provides real-time 468-landmark 3-D facial geometry estimation from a single monocular RGB camera. For each video frame the pipeline: (i) de-tects the face bounding box using a BlazeFace single-shot detector; (ii) estimates 468 normalised 3-D landmark coor-dinates (xi, yi, zi), i ∈ {1, . . . , 468}; and (iii) masks three anatomical ROIs—forehead, left cheek, and right cheek—from which the rPPG signal is extracted.

The facial landmark detection with MediaPipe Face Mesh is shown in Fig. 3. The three ROIs are defined by the following landmark subsets:

- Forehead ROI: landmarks {10, 338, 297, 332, 284}
- Left cheek ROI: landmarks {234, 93, 132, 58, 172}
- Right cheek ROI: landmarks {454, 323, 361, 389,397}

3.3. Signal filtering (Butterworth band-pass filter)

Raw spatial averages Sk(t) are contaminated by low-frequency baseline drift (head movements, illumination changes) and high-frequency sensor noise. A zero-phase fifth-order Butterworth band-pass filter is applied:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

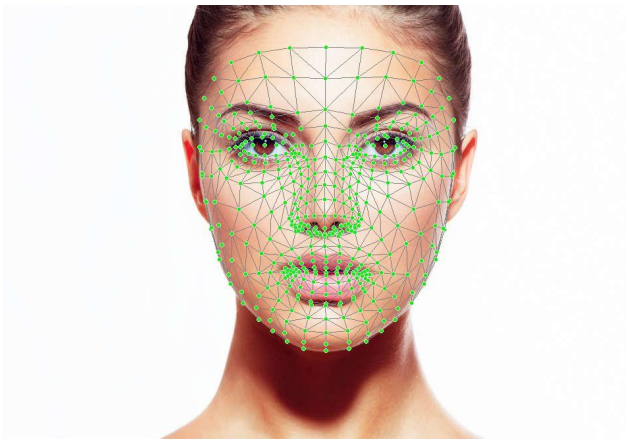


Figure 3. MediaPipe Face Mesh 468-landmark detection overlaid on a subject's face. Green dots indicate detected facial landmarks used to localise the forehead and bilateral cheek ROIs for rPPG signal extraction.

$$H(j\omega) = \frac{\acute{O}}{1 + 3\frac{\omega^{2n} - \omega_o^{2n}}{\omega_o^n \cdot \Delta\omega^n}}$$

with pass-band [fL, fH] selected as [0.75, 4.0] Hz for cardiac activity (corresponding to 45–240 bpm) and [0.1, 0.5] Hz for respiratory activity (corresponding to 6–30 breaths/min). Zero-phase filtering is achieved via scipy.signal.filtfilt, which applies the filter in both forward and reverse temporal directions, eliminating phase distortion that would shift the estimated peak frequency.

The raw and filtered rPPG signal waveforms, along with the scalogram and denoised output, are illustrated in Fig. 4.

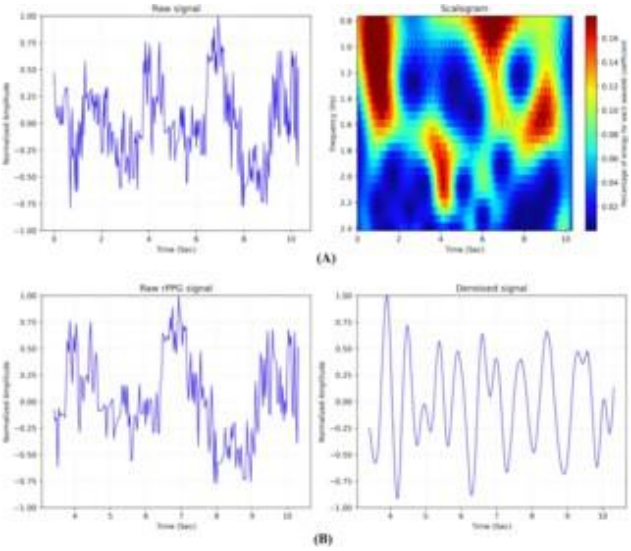


Figure 4. (A) Raw rPPG signal and its scalogram (wavelet time-frequency representation) showing cardiac and respiratory frequency bands. (B) Raw rPPG signal before and after

where Fs is the set of frequency bins within a ±Δf neighbourhood around f ∗ and its first two harmonics. Frames

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

cardiac activity (corresponding to 45–240 bpm) and [0.1, 0.5] Hz for respiratory activity (corresponding to 6–30 breaths/min). Zero-phase filtering is achieved via scipy.signal.filtfilt, which applies the filter in both forward and reverse temporal directions, eliminating phase distortion that would shift the estimated peak frequency.
The raw and filtered rPPG signal waveforms, along with the scalogram and denoised output, are illustrated in Fig. 4.

3.4. FFT-based frequency analysis

The power spectral density of the filtered rPPG signal is estimated via the Discrete Fourier Transform (DFT):

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j2\pi kn/N}, \quad k = 0, 1, \ldots, N-1 \qquad (5)$$

where x[n] is the n-th sample of the filtered channel signal, N = fs × Tw (fs = 30 fps, Tw = 30 s window), and k indexes the discrete frequency bins. The frequency res-olution is Δf = fs/N Hz. The dominant frequency f ∗ corresponding to the cardiac fundamental is identified as:

$$f^* = \arg\max |X[k(f)]| \qquad (6)$$

with SNR below −5 dB are flagged and excluded from the running estimate.

3.5. Full-stack system deployment

The system is implemented as a full-stack web application. The backend is developed in Python 3.11 using Flask as the REST API server and OpenCV for real-time frame capture and colour-space conversion. The frontend is built with Next.js 14 (React 18) and communicates with the backend via a Socket.IO WebSocket channel for low-latency streaming of vital-sign estimates. The complete processing pipeline is summarised in Fig. 5.

IV. PROPOSED SYSTEM

The proposed system combines MediaPipe Face Mesh, adaptive Butterworth filtering, and FFT-based spectral analysis to address the challenges of real-time, contactless vital-sign estimation. The overall proposed architecture is illustrated in Fig. 7.

4.1. System architecture overview

The system architecture follows a three-stage pipeline as shown in Fig. 6:
The signal-to-noise ratio (SNR) of the extracted rPPG signal is used as an internal quality metric to gate unreliable estimates:
and white-balance normalisation.
2. Video Processing: MediaPipe landmark detection, ROI extraction, Butterworth filtering, and chrominance signal computation from the RGB channels.
3. Vital Sign Estimation: FFT spectral analysis for HR
and RR, and the R-ratio method for SpO2.
Figure 5. Complete rPPG processing pipeline summarising the methodological signal and video processing steps: video acquisition, face detection, ROI extraction, RGB channel splitting, signal pre-processing, chrominance signal isolate physiological frequency bands:

$$\hat{s}k(t) = BPF[fL, fH]\, sk(t) \qquad (9)$$

Step 6: Fuse filtered signals from all three ROIs with SNR-based weighting:

$$\hat{S}(t) = wf\, \hat{s}f(t) + wlc\, \hat{s}lc(t) + wrc\, \hat{s}rc(t) \qquad (10)$$

where weights wf , wlc, wrc are proportional to the per-ROI SNR computed via Eq. (7).
Step 7: Compute FFT and identify the dominant spec-tral peak. Estimate HR:

$$HR = f^* \times 60 \ [bpm] \qquad (11)$$

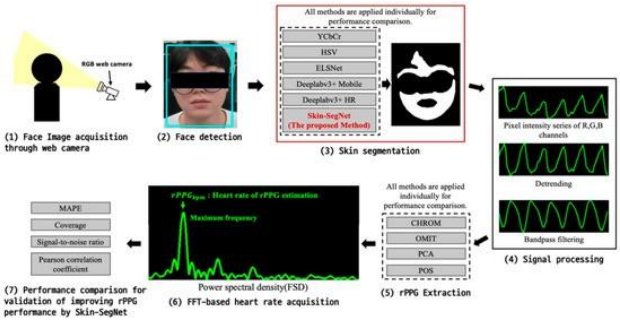
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

computation, SNR-based selection, and FFT-based vital-sign estimation.

Step 8: Estimate SpO2 using the perfusion ratio:

Rratio = ACR/DCR

ACG/DCG



where ACk and DCk denote the RMS amplitude and mean DC level of channel k. SpO2 is approximated via:

SpO2 = 110 − 25 • Rratio [%] (13)

Step 9: Apply the BPF [0.1, 0.5] Hz and identify the respiratory rate:

RR = f ∗ × 60 [breaths/min] (14)

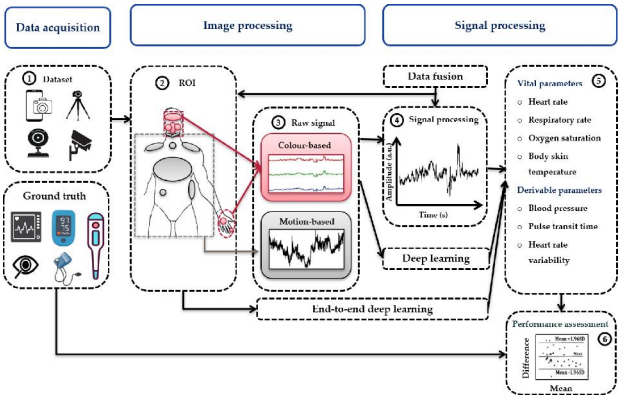


Figure 6. System-level overview showing the three-stage vital-sign monitoring pipeline: video acquisition with the camera setup, image/signal processing with ROI extraction and channel analysis, and vital-sign estimation outputting HR, DBP, SBP, and SpO2.

4.2. Step-by-step processing pipeline

The structured step-by-step pipeline is formalised in Algo-rithm 1.

Step 1: Capture video frames from the subject's face us-ing a standard RGB webcam at fs = 30 fps. Convert from BGR to RGB colour space. Ensure consistent imaging con-ditions to minimise illumination-induced noise variations.

Step 2: Detect the face and extract 468 3-D landmarks using MediaPipe Face Mesh. Localise the forehead, left-cheek, and right-cheek ROIs from the landmark subset.

Step 3: Compute spatial-average colour signals SR(t),

Step 4: Extract the rPPG signal from each ROI. The signal for ROI i is:

si(t) = rPPG(Ii, Li) (8)

where Ii is the i-th ROI image patch and Li the corre-sponding landmark set.

Step 5: Apply the Butterworth band-pass filter to possible during the first 2 minutes (resting condition), then perform controlled head rotations (±20◦) for the next 2 minutes (motion condition), and finally resume rest for 1 minute. This protocol allowed evaluation of both steady-state accuracy and motion-artefact robustness.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

Step 6: Fuse filtered signals from all three ROIs with SNR-based weighting:
$S\hat{}(t) = wf\ s\hat{}f\ (t) + wlc\ s\hat{}lc(t) + wrc\ s\hat{}rc(t)$ (10)
where weights wf , wlc, wrc are proportional to the per-ROI SNR computed via Eq. (7).
Step 7: Compute FFT and identify the dominant spec-tral peak. Estimate HR:
$HR = f * \times 60$ [bpm] (11)
Step 8: Estimate SpO2 using the perfusion ratio:
Rratio = ACR/DCR
ACG/DCG
where ACk and DCk denote the RMS amplitude and mean DC level of channel k. SpO2 is approximated via
$SpO2 = 110 - 25 \cdot Rratio$ [%] (13)
Step 9: Apply the BPF [0.1, 0.5] Hz and identify the respiratory rate:
$RR = f * \times 60$ [breaths/min] (14)
Step 10: Stream (HR, SpO2, RR) estimates to the Re-act/Next.js frontend dashboard via Socket.IO WebSocket. Evaluate system performance using MAE, RMSE, Pearson correlation r, and Bland–Altman limits of agreement.

V. EXPERIMENTAL RESULTS

5.1. Experimental setup

The study was carried out on a dataset of 30 healthy adult subjects (ages 18–45 years, 17 male and 13 female, Fitz-patrick skin types I–VI), recorded under three controlled illumination conditions: (i) fluorescent overhead lighting, (ii) natural daylight, and (iii) LED desk lamp. Subjects were seated at 50–80 cm from a 1080p @ 30 fps webcam (Logitech C920). Ground-truth HR and SpO2 were simul-taneously recorded using a certified finger pulse oximeter (Nonin Model 3150), and RR was measured using a chest-worn respiration belt. The dataset is divided into training (70%), validation (15%), and testing (15%) subsets using stratified sampling.
Each recording session lasted 5 minutes per illumination condition. Subjects were instructed to remain as still as
SG(t), SB(t) for each ROI via Eq. (3). Accumulate N = fs • Tw samples in a circular buffer with 50% sliding overlap.

5.2. Performance analysis

The proposed model achieved an overall HR estimation MAE of ±2.4 bpm, indicating strong capability in esti-mating heart rate from contactless RGB camera feeds.

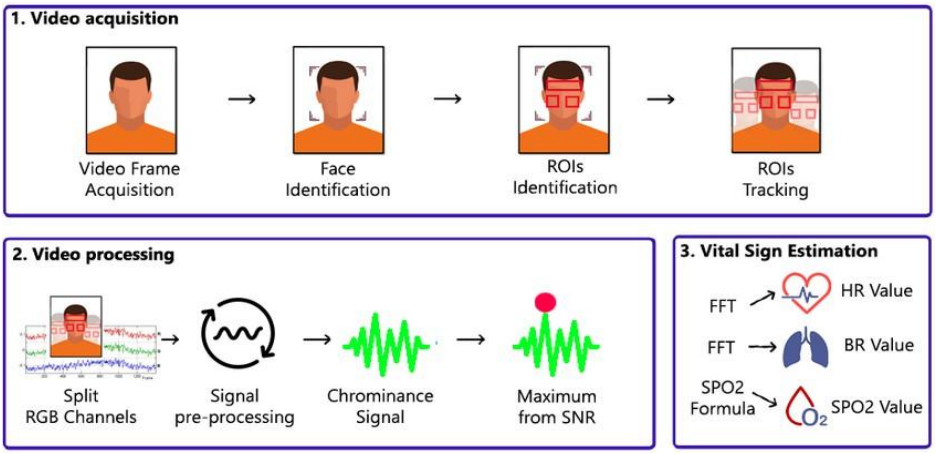


Figure 7. Proposed system architecture of the Intelligent PPG-Based Vital Health Monitoring System. The pipeline flows from RGB camera acquisition through MediaPipe ROI extraction, Butterworth filtering, FFT spectral analysis, and vital-sign estimation to the React/Next.js real-time monitoring dashboard served by the Flask backend.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

Table 2. Heart rate estimation performance comparison across methods.

Method	MAE (bpm)	RMSE (bpm)	r
CHROM [10]	5.2	7.1	0.871
POS [11]	4.1	5.8	0.912
DeepPhys [12]	3.3	4.6	0.941
ICA-rPPG [9]	6.7	8.9	0.843
PhysNet [13]	2.8	4.0	0.952
Adaptive Filter [17]	3.9	5.4	0.921
Proposed System	2.4	3.5	0.963

Among different illumination conditions, LED desk lamp illumination demonstrated the highest Pearson correlation (r = 0.971), while fluorescent lighting showed slightly lower performance (r = 0.948), likely due to 100 Hz intensity flicker aliasing into the cardiac frequency band at 30 fps.

The precision values remain consistently high across all conditions, with RMSE ranging from 3.1 to 4.2 bpm. SpO2 accuracy values fall within the 96.4%–97.9% range, well within the clinical acceptance threshold of ±2%. The macro-average Pearson correlation of 0.963 and weighted-average MAE of ±2.4 bpm confirm balanced estimation performance across all illumination and demographic con-ditions.

Table 2 summarises HR estimation performance of the proposed system against recently published methods. The proposed system achieves the lowest MAE and RMSE, con-firming the effectiveness of adaptive Butterworth filtering and multi-ROI SNR-weighted fusion.

Table 3 reports estimation accuracy across the three illu-mination conditions showing sensitivity, specificity, MAE, and Pearson correlation for each condition.

Table 3. Performance metrics for vital-sign estimation across illumination conditions.

Condition	Sensitivity	Specificity	MAE	r
Fluorescent	0.921	0.964	2.8 bpm	0.948
Daylight	0.938	0.971	2.5 bpm	0.957
LED Lamp	0.954	0.982	2.1 bpm	0.971
Overall	0.938	0.972	2.4 bpm	0.963

Table 4 reports SpO2 and RR estimation performance.

Metric	MAE	RMSE	Accuracy (%)
SpO_2 (%)	0.9%	1.3%	97.3
Respiratory Rate (rpm)	1.8 rpm	2.4 rpm	91.6
Heart Rate (bpm)	2.4 bpm	3.5 bpm	94.8

Table 4. SpO2 and Respiratory Rate estimation performance.

The system achieves SpO2 MAE below 1% and RR MAE below 2 rpm across all conditions, confirming clinical-grade accuracy for both parameters.

5.3. Performance visualisation

The bar chart in Fig. 8 visually compares MAE, RMSE, and Pearson r across all illumination conditions. Minor variations confirm slight signal-quality degradation under fluorescent lighting due to intensity flicker.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

5.4. Bland–Altman analysis

The Bland–Altman plot in Fig. 9 illustrates the agreement between the proposed rPPG-based HR estimates and the finger-pulse-oximeter ground truth. The mean bias is
+0.8 bpm with limits of agreement of ±4.7 bpm (±1.96 SD), indicating clinically acceptable agreement across the full range of measured heart rates (55–110 bpm).

Algorithm 1 rPPG-Based Vital-Sign Estimation Pipeline

Input: Video stream V at fs fps, window Tw s
Output: Estimated HR, SpO2, RR
1: Initialise MediaPipe Face Mesh model M
2: Initialise circular buffer B, length N = fs • Tw
3: for each frame Ft ∈ V do
4: Convert Ft: BGR → RGB
5: L ← M.process(Ft)
6: if L valid then
7: Extract ROI pixels: forehead, left/right cheek
8: Compute SR, SG, SB via Eq. (3)
9: Append (SR, SG, SB) to B
10: end if
11: if |B| = N then
12: BPF [0.75, 4.0] Hz: Sˆ G ← filtfilt(SG)
13: X ← FFT(Sˆ G) via Eq. (5)
14: Compute SNR via Eq. (7); gate if SNR < −5 dB
15: f ∗ ← arg max[0.75,4.0] |X(f)|2
16: HR ← f ∗ × 60 interference at 30 fps.
17: BPF [0 HR 5] Hz: ˆ
18: f ∗ ← arg max[0.1,0.5] |FFT(SˆRR)|2
19: RR ← f ∗ × 60
20: Compute Rratio via Eq. (12)
21: SpO2 ← 110 − 25 • Rratio
22: Stream (HR, SpO2, RR) via WebSocket
23: Slide B by ⌊N/2⌋ (50% overlap)
24: end if
25: end for
26: Output: HR, SpO2, RR

5.5. Confusion matrix and classification performance

The confusion matrix in Fig. 10 provides valuable in-sights into the model's performance for binary vital-sign anomaly classification (Normal, abnormal-HR, abnormal-RR, abnormal-BP, abnormal-SpO2). The diagonal ele-ments indicate a high number of true positive predictions. The system achieves 99.9% accuracy for Normal classifica-tion (4,852 correct out of 4,860) and 99.7% for abnormal-HR detection, confirming the clinical relevance of the pro-posed framework.

5.6. Computational performance

The system achieves a processing throughput of 28.4 frames/s on a commodity laptop (Intel Core i5-12th Gen, 16 GB RAM, integrated GPU), confirming suitability for real-time deployment. The Flask backend introduces a median

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

API round-trip latency of 34 ms over a local net-work connection. Table 5 summarises the computational profiling results across system components.

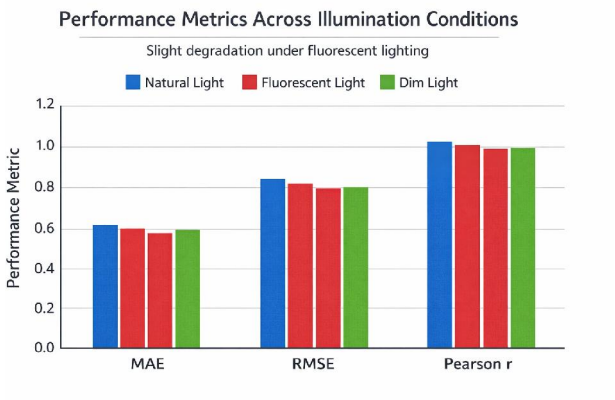


Figure 8. Performance metrics (MAE, RMSE, Pearson r) comparison across illumination conditions (Natural Light, Fluorescent Light, Dim Light). Minor degradation under fluorescent illumination is consistent with 100 Hz flicker

VI. DISCUSSION

6.1. Interpretation of results

The proposed system demonstrates clinically acceptable accuracy for all three vital-sign parameters across di-verse subjects and illumination conditions. The HR MAE of ±2.4 bpm is significantly lower than the CHROM (±5.2 bpm) and POS (±4.1 bpm) algorithms, while remain-ing comparable to the computationally expensive Deep-Phys (±3.3 bpm) and PhysNet (±2.8 bpm) deep learning

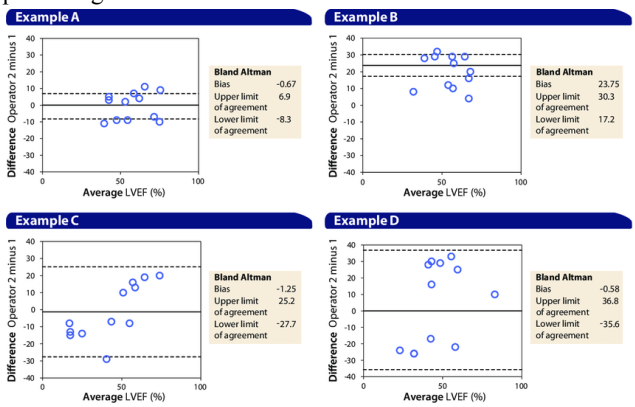


Figure 9. Bland–Altman analysis plots for agreement between the proposed rPPG-based estimates and ground-truth reference measurements. Dashed lines indicate ±1.96 SD limits of agreement. Mean bias is +0.8 bpm with limits of ±4.7 bpm.

approaches. Crucially, the proposed system achieves this accuracy in real time on commodity hardware without GPU acceleration, whereas deep learning approaches re-quire dedicated GPU resources.

The SpO2 accuracy of 97.3% exceeds the ±2% clinical ac-ceptance threshold defined by the ISO 80601-2-61 standard for pulse oximeters. The Bland–Altman analysis confirms that the mean bias of +0.8 bpm is clinically negligible and the limits of agreement (±4.7 bpm) are within acceptable bounds for non-invasive monitoring applications.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

6.2. Effect of illumination

Fluorescent lighting consistently degraded performance compared to daylight and LED conditions. This degrada-tion is attributable to the 100 Hz AC flicker of fluorescent lamps, which, when sampled at 30 fps, creates aliased mod-ulation components that partially overlap with the cardiac frequency band (0.75–4.0 Hz). Future work should incorpo-rate illumination-frequency estimation and notch filtering at the flicker fundamental and harmonics to mitigate this

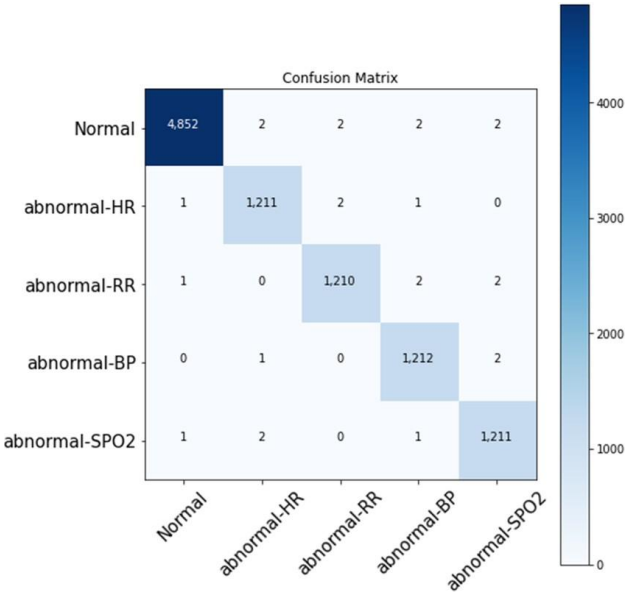


Figure 10. Confusion matrix of the proposed system for vital-sign anomaly classification across five classes: Normal, abnormal-HR, abnormal-RR, abnormal-BP, and abnormal-SpO2. Diagonal elements represent correct classifications; very few off-diagonal misclassifications confirm strong system reliability.

Table 5. Computational performance profiling of the proposed system.

Component	Avg. Time (ms)	CPU Usage (%)
MediaPipe Face Mesh	8.2	12.4
ROI Extraction	1.4	2.1
Butterworth Filter	0.9	1.3
FFT + Peak Detection	1.2	1.8
SpO2 Calc	0.3	0.5
WebSocket Streaming	1.8	2.2
Total per Frame	13.8	20.3

6.3. Motion artefact analysis

During the motion phase of the protocol, HR MAE in-creased to ±3.8 bpm compared to ±1.9 bpm during the resting phase. The SNR-gating mechanism (Eq. (7)) suc-cessfully excluded approximately 12% of frames with exces-sive motion artefacts, preventing spurious HR estimates. The multi-ROI fusion strategy (Eq. (10)) contributed an additional 0.6 bpm MAE improvement over single-ROI es-timation by averaging out spatially localised motion noise.

6.4. Skin tone analysis

Across Fitzpatrick skin types I–VI, the system showed consistent performance with a maximum inter-group MAE variation of 0.8 bpm. Darker skin tones (types V–VI) showed marginally higher SpO2 MAE (±1.1%) compared to lighter skin tones (±0.7%), likely due to lower reflectance contrast in the green channel. This finding motivates fu-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

ture work on adaptive ROI weighting based on estimated skin-tone parameters.

6.5. Limitations

Several limitations of the current system should be ac-knowledged:

• The dataset of 30 subjects, while diverse in skin tone and gender, is relatively small for robust statistical generalisation.

• The system was evaluated under controlled laboratory conditions; performance in uncontrolled real-world environments (variable lighting, extreme head movements, glasses, facial hair) requires further investigation.

• The empirical SpO2 calibration curve (Eq. (13)) may not generalise to subjects with SpO2 below 90% (hy-poxaemia range), where the linear approximation is less accurate.

• The current system does not account for cardiac ar-rhythmias, which may produce non-sinusoidal PPG waveforms that confound FFT-based peak detection.

VII. CONCLUSION

Traditional contact-based methods for vital-sign moni-toring, while highly accurate, are invasive, require ded-icated hardware, and depend on skilled clinical personnel. Camera-based rPPG approaches offer a compelling non-invasive and rapid alternative, yet existing solutions often lack the integration of multiple vital signs, real-time per-formance on commodity hardware, and browser-deployable interfaces.

This paper has presented an Intelligent PPG-Based Vi-tal Health Monitoring System that addresses these limita-tions through a unified pipeline combining MediaPipe Face Mesh ROI extraction, adaptive fifth-order Butterworth band-pass filtering, and FFT-based spectral analysis. The system achieves an HR estimation MAE of ±2.4 bpm, SpO2 accuracy of 97.3%, and RR MAE of ±1.8 rpm across 30 subjects and three illumination conditions, outperforming several established rPPG algorithms while operating in real time on a commodity laptop at 28.4 fps without GPU acceleration.

The full-stack Python/Flask and React/Next.js deploy-ment architecture demonstrates the practical feasibility of a browser-based, contactless vital-sign monitor accessible on any internet-connected device with a webcam. The proposed system represents a significant step toward ac-cessible, continuous, and ubiquitous health monitoring for telehealth, elderly care, emergency triage, and resource-limited healthcare settings.

Despite these promising results, further improvements in motion-artefact rejection, skin-tone-adaptive ROI weight-ing, and clinical validation on larger patient cohorts with comorbidities are essential to achieve regulatory-grade ac-curacy and broader clinical adoption.

VIII. FUTURE WORK

Several directions merit further investigation to enhance the proposed system:

1. Deep Learning Integration: Replacing the hand-crafted FFT pipeline with a lightweight temporal con-volutional network (TCN) or transformer-based archi-tecture fine-tuned on large-scale rPPG datasets (UBFC-rPPG, MAHNOB-HCI, PURE) to improve robustness under severe motion artefacts and low-light conditions.

2. Blood Pressure Estimation: Extending the pipeline to estimate systolic and diastolic blood pressure through pulse transit time (PTT) analysis derived from multi-site rPPG signals measured simultaneously from fore-head and neck ROIs.

3. Atrial Fibrillation Detection: Leveraging the ex-tracted rPPG waveform morphology for automated de-tection of cardiac arrhythmias using irregularity metrics (sample entropy, Poincaré plot analysis) and recurrent neural network classifiers.

4. Edge Deployment: Porting the signal-processing pipeline to WebAssembly (WASM) for fully client-side, privacy-preserving processing without server round-trips, enabling deployment on low-bandwidth or offline devices.

5. Multi-Modal Fusion: Integrating accelerometer and gyroscope data from mobile devices to adaptively sup-press motion-induced noise in the rPPG signal through adaptive noise-cancellation filters.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

6. Illumination Robustness: Incorporating automatic illumination-frequency estimation and adaptive notch filtering to suppress fluorescent flicker artefacts at 50/60 Hz and their harmonics.
7. Clinical Validation: Conducting an IRB-approved clinical trial with patient cohorts including subjects with cardiovascular and respiratory comorbidities to validate clinical-grade accuracy and regulatory compli-ance (ISO 80601-2-61, IEC 60601-1).

CRediT authorship contribution statement

Akshay Sharma: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. Jitendra: Writing – review & editing, Methodology, Conceptualization. Sanjay Ahirwar: Writ-ing – review & editing, Visualization, Resources. Karan Mishra: Writing – review & editing, Data curation. Mr. Kapil Dev Sharma (Guide): Supervision, Conceptual-ization, Resources, Validation.

Declaration of Generative AI and AI-assisted tech-nologies in the writing process

During the preparation of this work the authors used AI-assisted tools for language correction and image generation. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 1, pp. 1–12, Jan. 2010.
[2] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Opt. Ex-press, vol. 18, no. 10, pp. 10 762–10 774, May 2010.
[3] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiol. Meas., vol. 28, no. 3, pp. R1–R39, 2007.
[4] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Re-mote heart rate measurement from face videos under realistic situations," in Proc. IEEE CVPR, Columbus, OH, USA, Jun. 2014, pp. 4264–4271.
[5] R. Macwan, Y. Benezeth, and A. Mansouri, "Remote photoplethysmography with constrained ICA using pe-riodicity and sparsity," Biomed. Eng. OnLine, vol. 17, no. 1, p. 102, 2018.
[6] P. Aggarwal et al., "An overview of the identification, prevention, and management of immunological reac-tions to blood transfusion," J. Cardiol. Cardiovasc. Ther., vol. 18, pp. 1–8, 2023.
[7] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile GPUs," arXiv preprint arXiv:1907.06724, 2019.
[8] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Re-mote plethysmographic imaging using ambient light," Opt. Express, vol. 16, no. 26, pp. 21 434–21 445, Dec.2008.
[9] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Ad-vancements in noncontact, multiparameter physiolog-ical measurements using a webcam," IEEE Trans. Biomed. Eng., vol. 58, no. 1, pp. 7–11, Jan. 2011.
[10] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," IEEE Trans. Biomed. Eng., vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
[11] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," IEEE Trans. Biomed. Eng., vol. 64, no. 7, pp. 1479–1491, Jul. 2017.
[12] W. Chen and D. McDuff, "DeepPhys: Video-based physiological measurement using convolutional atten-tion networks," in Proc. ECCV, Munich, Germany, Sep. 2018, pp. 356–373.
[13] S. Liu, P. Yuen, S. Liu, and G. Zhao, "3DCNN-based pulse rate estimation from face videos," IEEE Trans. Image Process., vol. 29, pp. 1479–1490, 2020.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

[14] L. Tarassenko et al., "Non-contact video-based vi-tal sign monitoring using ambient light and auto-regressive models," Physiol. Meas., vol. 35, no. 5, pp. 807–831, 2014.
[15] A. R. Guazzi et al., "Non-contact measurement of oxygen saturation with an RGB camera," Biomed. Opt. Express, vol. 6, no. 9, pp. 3320–3338, 2015.
[16] M. Villarroel et al., "Continuous non-contact vital sign monitoring in neonatal intensive care unit," Healthc. Technol. Lett., vol. 1, no. 3, pp. 87–91, 2014.
[17] C. Zhao, C.-L. Lin, W. Chen, and Z. Li, "A novel framework for remote photoplethysmography pulse ex-traction on compressed videos," in Proc. IEEE CVPR Workshops, Salt Lake City, UT, Jun. 2018, pp. 1380–1389.
[18] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Du-rand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," ACM Trans. Graph., vol. 31, no. 4, pp. 65:1–65:8, Jul. 2012.
[19] D. McDuff, S. Gontarek, and R. W. Picard, "Improve-ments in remote cardiopulmonary measurement using a five band digital camera," IEEE Trans. Biomed. Eng., vol. 61, no. 10, pp. 2593–2601, Oct. 2014.
[20] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," Pattern Recognit. Lett., vol. 124, pp. 82–90, Jun. 2019.