IET Wireless Sensor Systems

WILEY

The Institution of Engineering and Technology

REVIEW OPEN ACCESS

Contactless Health Monitoring: An Overview of Video-Based Techniques Utilising Machine/Deep Learning

Alaa Hajr[1] | Bahram Tarvirdizadeh[1] | Khalil Alipour[1] | Mohammad Ghamari[2]

[1]Advanced Service Robots (ASR) Laboratory, Department of Mechatronics Engineering, College of Interdisciplinary Science and Technology, University of Tehran, Tehran, Iran | [2]Department of Electrical Engineering, California Polytechnic State University, San Luis Obispo, California, USA

Correspondence: Bahram Tarvirdizadeh (bahram@ut.ac.ir)

Received: 12 January 2025 | **Revised:** 3 May 2025 | **Accepted:** 19 May 2025

Handling Editor: Helsham El-Sayed

Funding: The authors received no specific funding for this work.

Keywords: biomedical measurement | biosensors | computer vision in healthcare | deep learning (DL) | machine learning (ML) | remote photoplethysmography (rPPG)

ABSTRACT

Vital signs are crucial indicators of an individual's physiological well-being and represent one of the primary evaluations conducted in clinical and hospital environments. A comprehensive evaluation of a patient's health state depends on these signs which include heart rate (HR), respiratory rate (RR), blood oxygen saturation (SpO2), blood pressure (BP) and body temperature (BT). In recent years, there has been significant interest in using imaging photoplethysmography (iPPG) with consumer-level cameras for contactless health monitoring (CHM) to accurately assess vital signs. The introduction of iPPG in CHM signifies the beginning of a remarkable era in the history of healthcare, whereby diagnostic processes are enhanced via the integration of technology and patient well-being. This review article presents a comprehensive examination of CHM techniques utilising machine learning (ML) and deep learning (DL) algorithms for the assessment of critical vital signs. The article addresses the challenges and research gaps identified in recent studies, particularly those related to variations in lighting conditions, head movements and the impact of different colour types on the accuracy and reliability of CHM techniques. Finally, we propose several recommendations aimed to enhance the efficiency of CHM systems. These include the development of more robust learning algorithms and the creation of diverse datasets that encompass a wide range of demographics including variations in gender, skin colour and lighting conditions.

1 | Introduction

Monitoring vital signs such as heart rate (HR), respiratory rate (RR), blood pressure (BP), body temperature (BT) and oxygen saturation (SpO2) is essential for assessing a patient's condition and guiding clinical decisions [1, 2]. Traditionally, patients directly place contact sensors such as temperature probes, blood pressure cuffs and ECG electrodes on their bodies to take these measurements. Although touch sensors are useful, they may cause pain and pose the danger of transmitting infectious illnesses. This is a significant challenge in clinical and hospital settings, where the problems of spreading germs across different areas and the need for rigorous sterilisation procedures are particularly tough for both healthcare personnel and patients [3, 4]. Many studies, such as those looking at the application of machine learning (ML) and deep learning (DL) techniques to photoplethysmography (PPG) signals, have produced effective methods for analysing PPG signals and extracting important health-related data within this framework. These studies have demonstrated that PPG signals,

obtained from sensors connected to the body, can reliably determine vital signs.

The key distinctions between traditional contact-based monitoring methods and video monitoring (VM)-based health monitoring are summarised in Table 1 which highlight the differences in accuracy, reliability and real-world applicability under various environmental conditions [5]. Although contact-based systems remain the clinical standard, video monitoring (VM)-based approaches offer significant advantages in terms of non-invasiveness, continuous monitoring and real-world feasibility [6, 7]. With ongoing advancements in AI-driven signal enhancement, VM-based systems are becoming increasingly viable for home healthcare, telemedicine and large-scale health assessments, marking a shift towards more adaptable and patient-friendly health monitoring solutions [7].

We estimated BP in our previous work using methods including random forest (RF), decision trees (DT), linear regression (LR) and XGB To estimate vascular age [8], we also used the k-nearest neighbour (KNN), DT and support vector machines (SVM) algorithms in our previous study [9]. Furthermore, we employed enhanced long short-term memory (LSTMs) [10, 11], a convolutional neural network (CNN) combined with multilayer perceptron (MLP) algorithm [12], in our next investigations to accurately detect stress. All these ideas, nevertheless, are considered to be affixed to the body's skin.

To mitigate the issues associated with contact sensors, there has been significant advancement in imaging photoplethysmography (iPPG) techniques. iPPG is an innovative method that uses RGB cameras because of their widespread availability and popularity among the general public. This approach eliminates the need for physical contact, promoting non-invasive diagnostics and enhancing patient comfort and safety [3, 13]. In previous studies, this signal has been referred to by various names including iPPG, PPGi, PPGI, camera-based PPG, remote PPG (rPPG), distance PPG, non-contact PPG or video PPG [9]. In this paper, we use iPPG to refer to all image-based PPG measures.

iPPG measures subtle skin colour variations or cyclic movements of the body to estimate vital signs. The iPPG signal is extracted using a light source and a photodetector, such as an RGB camera, to quantify variations in the diameters of blood vessels beneath the skin [6, 14]. The iPPG signal has been generated by the photodetector's detection of subtle variations in light intensity resulting from blood circulation, either via reflection or transmission. The concept is derived from the Beer–Lambert law, which establishes a direct correlation between the extent of light penetration, the concentration of haemoglobin and the quantity of light absorbed by the blood as seen in Figure 1 [8].

Contactless health monitoring (CHM) offers significant advantages in patient safety and comfort while addressing the limitations of traditional sensor-based methods [10, 18, 19]. Table 2 highlights the rapid growth of research in CHM using video-based ML/DL techniques. These non-invasive approaches enhance measurement accuracy through advanced algorithms and effectively handle challenges such as lighting variability,

TABLE 1 | Performance comparison of contact-based and video-based health monitoring approaches [5].

Criterion	Contact-based methods	VM-based health monitoring
Accuracy	High but affected by sensor misplacement & motion artefacts.	Sensitive to lighting and motion but continuously improving
Reliability	Reliable in clinical settings but uncomfortable for prolonged use.	Enhancing with ML/DL; still affected by motion and illumination variations.
Real-world applicability	Limited to clinical settings, hygiene concerns.	Ideal for remote monitoring; requires optimisation for real-world deployment.

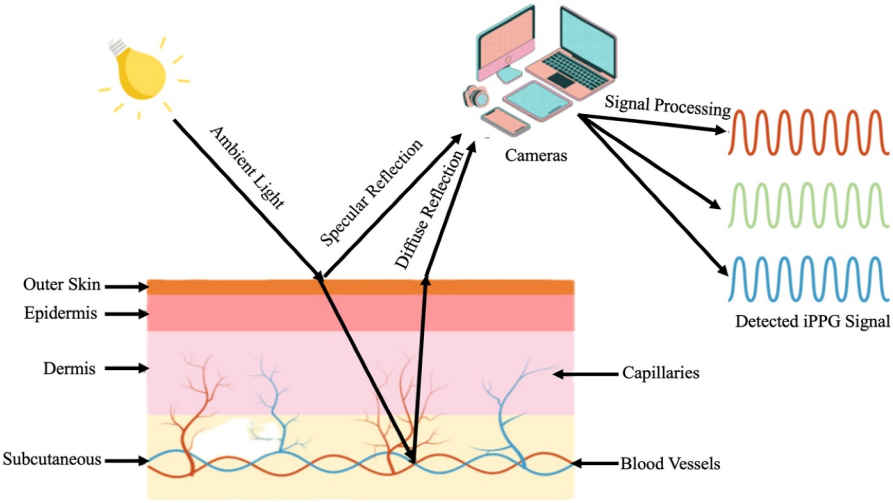

FIGURE 1 | Illustrates the process of extracting the iPPG signal from the three channels (red–green–blue) to acquire vital signs [15–17].

TABLE 2 | Comparative analysis of review articles on CHM utilising video-based techniques.

Ref. and year	Major contribution	Covered topics
[12] 2021	DL methods for HR measurement.	HR
[6] 2021	Include: history and background, technical details, recent advances, progress challenges and directions for future work.	Non-contact physiologic monitoring
[13] 2022	Overview of the physiological and technical aspects behind the estimation of vital signs	Vital signs
[20] 2022	ML models and videos of facial regions for estimating HR.	HR
[14] 2022	Aims to answer the following research questions: 1. Which vital sign is monitored using what type of camera? 2. What is the performance and which factors affect it? 3. Which health issues are addressed by camera-based techniques?	BVP, HR, RR, SpO2, and BP
[21] 2023	Visual contactless physiological monitoring (VCPM) is capable of using videos based on DL.	BVP, HR, RR, SpO2, and BP
[18] 2023	Utilising a camera to capture and analyse small colour fluctuations on the surface of the human skin in order to collect iPPG data.	HR, RR, and SpO2
[22] 2024	ML and DL for the healthcare system, concentrating on vital state-of-the-art features, integration benefits, applications and future guidelines.	Smart healthcare
[19] 2024	A comprehensive review of traditional computer vision and DL-based methods for estimating HR using contactless.	HR
[5] 2024	A framework for vital signs estimation using video analysis and ML.	Vital signs
[23] 2024	Vital signs estimation based on DL.	HR, BP, and SpO2
[24] 2024	Presents iPPG measurement using computer vision and DL for HR estimation.	HR with DL
This review	This comprehensive review explores the application of ML and DL in video-based vital sign estimation, emphasising enhancements in measurement accuracy and addressing prevalent technical challenges.	Vital signs with ML and DL. Challenges and limitation in video-based monitoring.

Abbreviation: BVP, Blood Volume Pulse.

motion artefacts and skin tone diversity. As a result, CHM technologies are transforming patient care by enabling continuous, remote monitoring across various real-world settings.

To better understand recent advancements in the field of study in recent years (as shown in Figure 2), This paper examines previous studies that have focused on vision-based methods and provides comparisons based on four significant factors: (1) The development of ML and DL algorithms for the iPPG signal; (2) the types and specifications of cameras; (3) the use of empirical datasets and (4) the accuracy of vital sign monitoring, measured by a variety of criteria such as MAE, RMSE and STD. In contrast to other review studies (presented in Table 2), this paper analyses and classifies vision-based methods that employ various learning algorithms such as ML and DL. This review identifies the ongoing limitations in hardware and software related to the assessment and measurement of each vital sign. In order to improve the reliability and accuracy of CHM in real clinical situations, we offer specific recommendations for future research.

Building on this foundation, the principal aim of this review is to outline significant research prospects and challenges, including novel algorithm development, video data-based hyperparameter learning, model enhancement and adaptation to devices with limited resources, as well as efficient data representation. A potential outcome of this endeavour is the formulation of a 'state-of-the-art ML/DL techniques' solution.

Consequently, this review aims to be used as a comprehensive resource for academics and industry specialists seeking to improve their knowledge and develop data-driven intelligent systems. The ultimate goal is to better the field of CHM by using innovative ML or DL techniques. Although this review does not follow the formal PRISMA framework, it adopts a structured narrative approach with a focused and methodologically-informed selection of recent literature relevant to video-based health monitoring. The main conclusions of this review may be expressed as follows:

1. Addressing the main limitations and difficulties in using video-based techniques to interact with CHM.
2. Enhancing the estimate of vital signs by improving the ML/DL techniques used in iPPG signal processing.
3. Understanding the benefits of enhancing CHM via the use of ML/DL techniques.

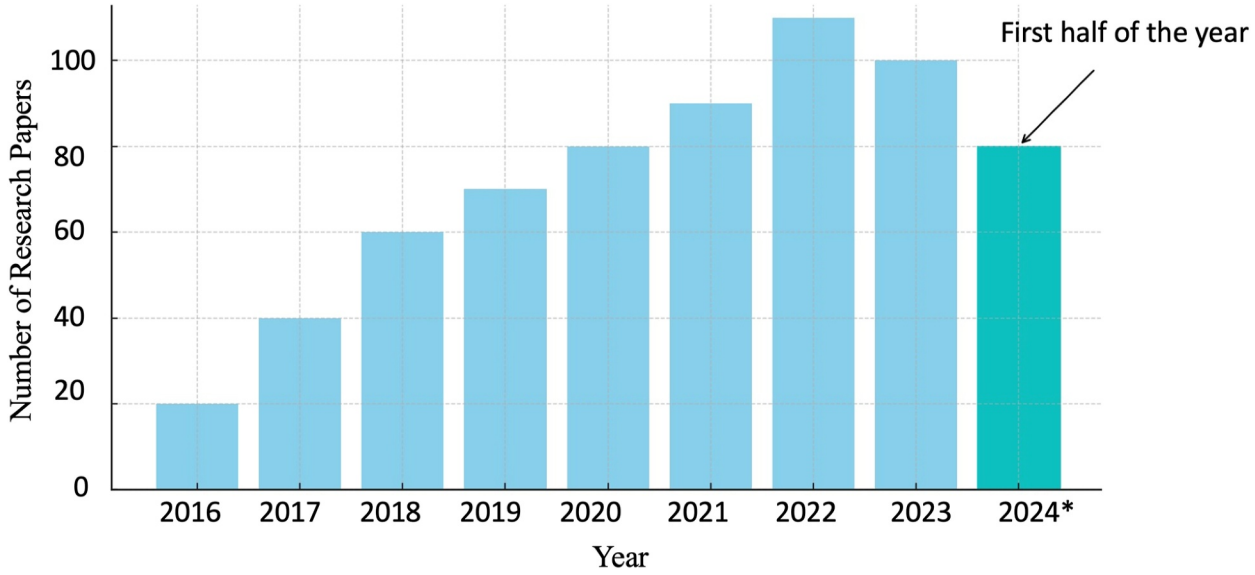

FIGURE 2 | Number of research articles using CHM with ML/DL (2016—June 2024).

4. Recommending CHM approaches that are particularly suitable for devices with restricted resources.

The review has been organised as follows: Section 2 discusses challenges in video-based vital signs monitoring, providing an overview of key obstacles in the field. Next, Section 3 covers various aspects of video monitoring (VM), focusing on how ML–DL techniques are utilised to track vital signs such as HR, HRV, RR and BP. Following this, Section 4 presents the conclusion, summarising key insights from the review. Finally, Section 5 explores future directions, highlighting potential advancements in video-based vital sign monitoring.

2 | Challenges in Video-Based Vital Signs Monitoring

2.1 | Impact of Environmental Factors on Accuracy

The accuracy and reliability of iPPG-based CHM systems are highly sensitive to variations in lighting conditions, head movements and skin tone differences. These environmental factors can introduce distortions, reduce signal quality and compromise clinical applicability. Addressing these challenges requires a combination of advanced pre-processing techniques and algorithmic improvements [25–27].

2.1.1 | Lighting Condition

Lighting variations significantly influence the intensity and absorption of light by the skin, leading to signal inconsistencies. In low-light environments, signal strength decreases, making it difficult to extract reliable physiological measurements, whereas overexposure results in saturation, distorting HR and RR estimations. Shadows and fluctuating illumination further exacerbate these inconsistencies, leading to reduced model accuracy and reliability [28, 29].

To mitigate these effects, contrast-limited adaptive histogram equalisation (CLAHE) and shading correction algorithms were integrated to enhance contrast and normalise brightness levels, improving feature extraction under variable lighting conditions [30]. Additionally, chrominance-based signal processing methods, such as CHROM and POS, were employed to extract pulse signals while minimising the impact of illumination interference [31]. Furthermore, independent component analysis (ICA) separates physiological signals from non-uniform lighting effects, improving robustness in real-world applications [32].

Recent advancements have specifically targeted rPPG extraction under extremely low-light environments. Xi et al. [33] proposed a low-light video enhancement method based on the Retinex theory to improve the visibility of facial blood flow signals before pulse extraction, significantly boosting rPPG performance in dim conditions. Additionally, a singular spectrum analysis and weighted combination framework was developed to reconstruct clean pulse signals even when traditional methods fail because of severe noise dominance [34]. More recently, an image enhancement model (IEM) integrated with deep learning-based and conventional rPPG pipelines demonstrated consistent improvements across multiple public datasets, validating its effectiveness for robust vital sign monitoring under low illumination [35].

2.1.2 | Motion Artefacts and Head Movements

Head movements and motion artefacts pose a significant challenge in real-world CHM applications. Unstable regions of interest (ROI) tracking because of movement can lead to misalignment and motion blur, ultimately affecting HR and BP estimations. These artefacts introduce signal distortions that reduce accuracy and hinder reliable physiological assessment [36, 37].

To address motion-induced errors, real-time facial tracking algorithms, including multi-task convolutional neural network

(MTCNN) [38], and RetinaFace [39], were implemented to ensure stable and consistent ROI tracking [5, 36]. In addition, dense optical flow and block matching algorithms compensate for motion distortions, aligning frames and reducing misalignment errors [37, 40]. DL approaches, particularly recurrent neural Nnetworks (RNNs) and long short-term memory (LSTM) models [38], enhance signal continuity and reconstruct missing data points, making the extracted physiological signals more reliable despite movement [31].

2.1.3 | Skin Tone Variations

Skin tone variability is another key factor affecting the accuracy of iPPG-based CHM systems. Differences in pigmentation influence light absorption and reflection, leading to inconsistencies in extracted physiological signals. Darker skin tones tend to absorb more light, reducing the intensity of reflected signals, whereas lighter skin tones reflect more light, leading to greater signal variation [28, 41].

Transforming the RGB colour space into YUV or HSV enhances signal extraction by reducing colour dependency, whereas CLAHE improves pulse detection uniformity across skin tones. Training DL models on diverse datasets minimises bias and enhances CHM generalisability [19]. Recent DL advancements, such as fully convolutional networks (FCNs) for automatic skin region extraction, improve feature selection and signal reliability, reducing reliance on traditional segmentation techniques and making iPPG processing more adaptive and robust [42].

2.2 | Clinical Significance of Performance Metrics

For iPPG-based CHM systems to be viable for medical applications, their accuracy and reliability must align with established clinical standards. Although DL and ML models have advanced the field, discrepancies remain when compared to benchmarked regulatory requirements.

Regulatory bodies such as the AAMI, BHS and FDA have defined acceptable error margins for vital sign estimation. BP must adhere to an error threshold of $\pm$ 5 mmHg, whereas SpO2 measurements require an accuracy level within 2%, and HR and HR variability (HRV) must demonstrate a correlation coefficient greater than 0.9 to be considered clinically reliable [43, 44].

Recent advancements in DL models have demonstrated varying levels of compliance with these standards. EfficientPhys, for example, has achieved high clinical accuracy, with a mean absolute error (MAE) of 1.14 bpm, RMSE of 1.81 bpm and a correlation coefficient (ρ) of 0.99, making it a strong candidate for real-world medical deployment [45]. Conversely, earlier models such as HR-CNN, which exhibited an MAE of 14.48 bpm and a correlation coefficient of 0.50, were deemed unsuitable for clinical use because of their high error margins [46, 47]. Similarly, in BP estimation, U-Net achieved clinically acceptable accuracy, with an MAE of 6.9 mmHg for systolic BP and 4.43 mmHg for diastolic BP, whereas CNN-based models exceeded the acceptable limits, reporting an MAE of 11.54 mmHg for systolic BP and 8.09 mmHg for diastolic BP [48, 49].

By explicitly aligning performance metrics with clinical benchmarks, this revised section ensures that the discussion extends beyond computational accuracy to real-world healthcare applicability [50]. The integration of pre-processing techniques to mitigate environmental factors, alongside performance comparisons with clinical standards, strengthens the validity of AI-driven CHM systems. These refinements significantly improve the clarity, coherence and scientific rigour of the manuscript, aligning it with high-impact medical applications [51].

3 | Video Monitoring (VM)

VM is a significant advancement in the field of computer–human interaction because it builds upon fundamental concepts to enable easy and unobtrusive tracking of essential physiological indicators. The above method integrates video technology with ML and DL models, allowing for continuous and real-time monitoring of health metrics. VM employs sophisticated algorithms and cameras to examine visual cues associated with physiological responses, providing comprehensive health assessments [5, 7].

Combining VM with ML/DL enhances the accuracy of analysing subtle facial or thoracic changes for vital sign estimation. SVM have been used to calculate HR [52], RF has been used to determine RR [5] and support vector regression (SVR) has been used to estimate BP [22]. These approaches are crucial to the development of data analysis techniques and algorithms, especially for the assessment of vital signs. Predicting vital signs is done using ML methodologies, as illustrated in Figure 3a, which shows the computational tools and various steps required.

VM systems typically utilise ambient light and RGB cameras to monitor regions of interest (ROI), such as the face, hands, and chest, detecting subtle variations in skin pigmentation and motion to generate iPPG waveforms. Traditional signal processing methods, including independent component analysis (ICA) [53], plane orthogonal to skin (POS) [16], Green channel information (Green) [54], and chrominance-based signals (CHROM) [55], have been widely employed for analysing these signals.

However, these conventional methods rely solely on rule-based signal processing techniques which fail to incorporate the biological and optical complexities of human skin imaging [56]. As a result, they are highly sensitive to motion disturbances, lighting variations and skin tone differences, leading to inconsistent and unreliable outcomes in real-world applications [56, 57].

By integrating ML and DL algorithms, CHM systems significantly enhance robustness and reliability in real-world applications by overcoming the limitations of traditional rule-based methods. Conventional techniques, such as ICA and POS, have demonstrated high sensitivity to motion artefacts, lighting variations and skin tone differences, often resulting in inconsistent

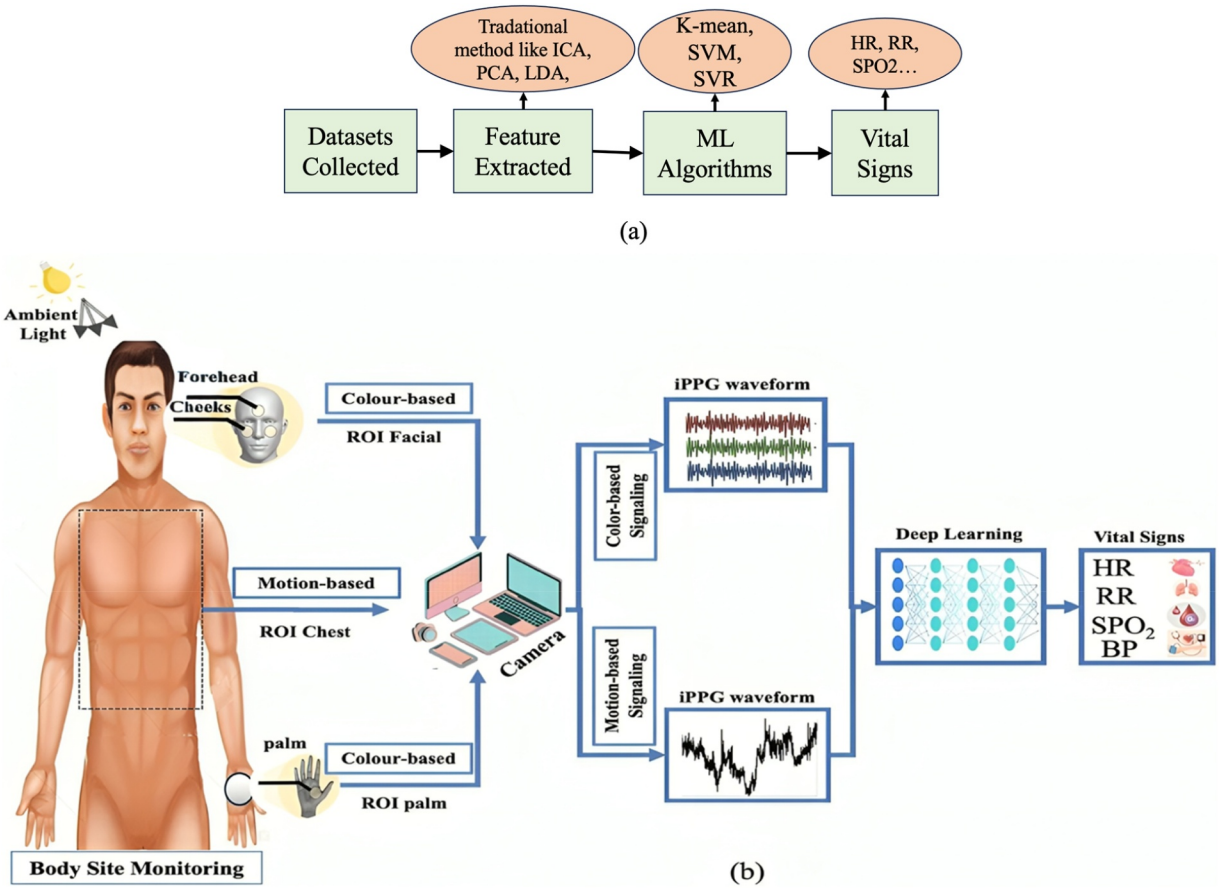

FIGURE 3 | (a) Basic block diagram of ML methods and (b) the camera is used to capture the iPPG signal from a specific body location.

performance, particularly in dynamic environments. Studies have shown that ICA achieves only 4% accuracy in high-motion scenarios, whereas PCA reaches 11%, and CHROM improves performance to 60% in moderate-motion conditions [5]. In contrast, DL models leverage spatial and temporal feature learning to dynamically adjust to environmental fluctuations, making them more resilient to noise and disturbances. For instance, EfficientPhys has demonstrated MAE of 1.14 bpm and an RMSE of 1.81 bpm, significantly outperforming traditional methods [5]. Similarly, DeepPhys achieved an MAE of 4.57 bpm, whereas *Meta*-rPPG reported an MAE of 3.01 bpm and RMSE of 3.68 bpm, indicating improved accuracy and consistency [5]. Notably, the dual-GAN model recorded the highest accuracy, with an MAE of 0.74 bpm, RMSE of 1.02 bpm and a correlation coefficient (ρ) of 0.997, highlighting the substantial advantage of DL in real-world physiological monitoring [58]. These results confirm that ML/DL-based approaches enhance signal stability, reduce error rates and improve adaptability to diverse conditions, making them far superior to conventional rule-based techniques for non-contact vital sign monitoring [5]. Figure 3b illustrates the process of body site monitoring using both colour-based and motion-based iPPG signal extraction, demonstrating the role of DL in vital sign estimation. In contrast to traditional methods that depend on established rules, contemporary DL models utilise spatial and temporal feature learning to improve signal extraction, rendering them more robust against noise and environmental fluctuations [59].

DeepPhys represents a significant advancement, introducing a CNN-RNN hybrid architecture to improve iPPG signal extraction [58]. DeepPhys utilises spatio-temporal patterns in visual footage, facilitating: enhanced motion compensation, markedly improving the accuracy of HR and RR estimation. Augmented respiratory monitoring, especially in identifying nuanced facial alterations and thoracic movements linked to respiration [58].

Likewise, PhysNet and 3D-CNN enhanced HR and HRV prediction by deriving high-fidelity iPPG signals from unprocessed video streams [60]. In contrast to 2D-CNN models that predominantly emphasise spatial features, PhysNet effectively captures both spatial and temporal fluctuations, rendering it very proficient in HRV spectral analysis [33]. These developments signify a paradigm shift in video-based health monitoring, moving from rule-based algorithms to adaptable DL frameworks. Figure 4 highlights the increasing adoption of various ML and DL models in video-based health monitoring, emphasising the dominance of CNN and its variations in this domain. In contrast to traditional methods, DL models adaptively respond to changes in environmental variables, rendering them more appropriate for practical application. Nonetheless, although ML and DL methodologies surpass conventional algorithms in controlled settings, their applicability to varied populations, skin types and lighting circumstances continues to pose a significant issue. Additional clinical validation and regulatory endorsement are necessary prior to the widespread use of these models in practical healthcare settings [38].

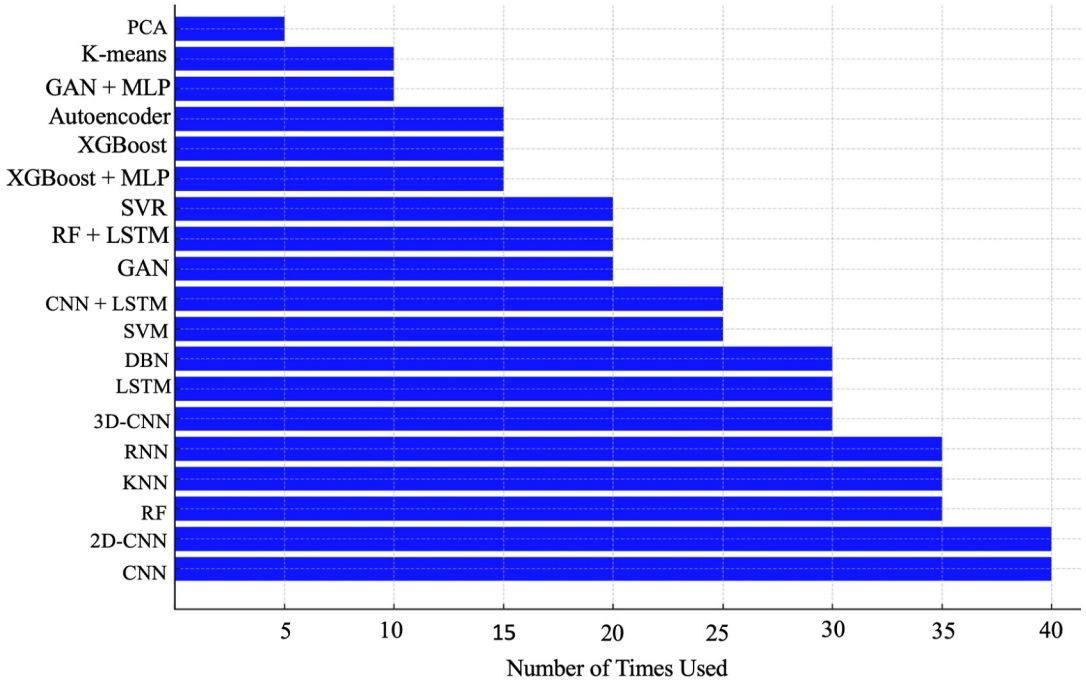

FIGURE 4 | Frequency of algorithm usage in CHM (2016—June 2024).

Although DL-based methods demonstrate superior accuracy and robustness compared to traditional model-based approaches, they often lack physical interpretability [61]. Unlike physiological and optical models that rely on well-understood biological principles, DL models operate as black-box systems, making it challenging to trace predictions back to underlying physiological phenomena [62]. This limitation raises concerns regarding the direct clinical applicability of DL-based rPPG systems. For safe and reliable deployment in healthcare settings, greater emphasis on model transparency and physiological grounding is required in future research [58].

VM employs a diverse range of cameras, such as thermal [63], infrared (IR) [64], RGB [34] and depth cameras [65]. The capability to accurately detect subtle variations in skin colour has resulted in the extensive utilisation of RGB cameras across several applications. In low-light conditions, IR cameras excel, whereas thermal cameras are ideal for monitoring temperatures. This study is focused mainly on articles that investigate RGB cameras, given their widespread popularity.

Notwithstanding these progressions, virtual reality has numerous technical obstacles including fluctuations in surrounding illumination. This issue is brought to attention in studies utilising the UBFC-rPPG dataset which was collected in authentic non-simulated environments [38]. Additionally, the MIHR database highlights the difficulties posed by lighting conditions ranging from 1 Lux to 100 Lux [66]. DL was utilised to successfully reduce the impact of motion disturbances [25, 67, 68]. In order to account for changes in skin tone, researchers employed four datasets that encompassed six distinct skin tones [69].

The MMPD dataset includes a diverse variety of skin tones, body movements and lighting conditions [70]. Its purpose is to improve the reliability and usefulness of iPPG technologies in real-world situations [41]. To further illustrate the characteristics of publicly available datasets used for video-based physiological signal analysis, Table 3 presents an overview of dataset specifications including the number of videos, dataset size, participant demographics and data collection conditions [52]. These datasets play a critical role in enhancing the generalisability and robustness of ML/DL models for VM applications. In order to enhance the precision of measurements, researchers are currently working on developing adaptive algorithms that can account for variations in lighting conditions, integrate motion compensation techniques and accommodate a wide range of skin tones. Future advancements in VM technology will focus on enhancing real-time processing capabilities and expanding its application across diverse healthcare domains. The continuous evolution of VM holds the potential to revolutionise patient monitoring and care delivery by enabling accurate, non-invasive and comprehensive health management.

3.1 | Heart Rate (HR) Detection Algorithms

Heart rate (HR), typically ranging from 60 to 100 bpm at rest, is a key indicator of cardiovascular health [82–85]. Abnormal HR, such as bradycardia or tachycardia, may signal underlying medical conditions [18, 42, 43, 85].

The accuracy of HR measurement using video-based techniques, which depend on hardware and algorithms, faces challenges such as lighting, motion and skin tone variability [56, 86–90]. Recent advancements have explored cost-effective and user-friendly solutions for contactless HR monitoring in home environments. For instance, Casalino et al. developed a smart mirror system using rPPG to estimate vital signs in real time which demonstrated improved robustness against abrupt head

TABLE 3 | Summary of public camera physiological measurement datasets.

Datasets	Total videos	Size (MB/GB)	Participants	Collection conditions
PURE 2014 [71]	60	42 GB	10 (2 F, 8 M)	Recorded in natural daylight with slight variations including six movement conditions: steady, talking, slow and fast horizontal movement and 20° (35 cm target) and 35° (70 cm target) rotation.
COHFACE 2016 [72]	160	386 MB	40 (12 F, 28 M)	4 videos per participant: good conditions (ceiling light + controlled blinds) and natural conditions (lights off).
BP4D+2016 [73]	1400	N/A	140 Majority female	Collected from individuals of diverse ethnic backgrounds, each performing 8 emotional tasks.
UBFC-rPPG 2017 [74]	42	75 GB	42 N/A	Varying amount of sunlight and indoor illumination
VIPL-HR 2019 [75]	2378	1000 GB	107 (79 F, 28 M)	Uncontrolled natural environments to simulate real-world conditions
OBF-HA 2020 [76]	212	N/A	200 (39 F, 61 M)	Collected from healthy individuals and atrial fibrillation patients in a controlled environment.
BH-rPPG 2020 [77]	105	N/A	35 (19 F, 16 M)	Under 3 illuminations: low, medium and high
MIHR 2020 [33]	165	~80 GB	15 (3 F, 12 M)	The dataset was recorded under eleven illumination levels (1–100 lux) in a darkroom using a fixed camera–subject setup.
UBFC-phys 2021 [78]	168	911 GB	56 N/A	The experiment included three stages to simulate social stress: Rest (T1), speech (T2) and mental arithmetic (T3).
MMSE-HR 2021 [73]	102	N/A	40 N/A	4 types of skin tones
SCAMPS 2022 [79]	2800	600 GB	2800 N/A	The synthetic videos were created through a precise simulation of blood flow and breathing effects using blender and 3D modelling techniques.
MMPD 2023 [70]	105	N/A	33 N/A	Vary in skin tone, motion and lighting conditions
Vital videos 2023 [80]	200	VV-small: 390 GB	100 (50% F, 50% M)	All skin tones were collected in various indoor locations across Western Europe, including libraries, retirement homes, shopping malls and community centres, ensuring diverse backgrounds and lighting conditions.
	500	VV-medium: 1000 GB	250 (50% F, 50% M)	
	1800	VV-large: 3200 GB	900 (50% F, 50% M)	
iBVP 2024 [81]	N/A	400 GB	33 (23 F, 10 M)	Participants performed four tasks: slow breathing, easy math, difficult math and guided head movement.

movements and achieved competitive results on the UBFC-rPPG dataset. This highlights the growing emphasis on integrating rPPG-based techniques into ambient assisted living frameworks and real-world applications.

Choosing an appropriate region of interest (ROI), such as the forehead, cheeks or neck, is essential because of consistent blood perfusion and visibility [66, 88, 91]. Several studies have used facial landmark detection tools (e.g., MediaPipe, MTCNN and RetinaFace) to isolate these areas for iPPG extraction [19, 92, 93].

ML and DL play a crucial role in improving the extraction of features required for reliable HR detection. For example, Ghanadian et al. [25] introduced SVM and k-nearest neighbour (KNN) algorithms; G. Antink et al. [94] proposed a Bayesian approach; Przybyło [19] utilised an LSTM-based method; and A. Anil et al. [39, 95] also employed LSTM architectures. Additionally, X. Niu et al. addressed the influence of skin tone variations in [87, 90].

Furthermore, maintaining consistent accuracy in varying operational contexts remains a persistent challenge. Adaptive filtering and advanced tracking algorithms have been developed to address these issues. Yet, the field continues to explore new methodologies to improve reliability and reduce latency in real-time applications [90, 96–99].

This section will detail the various algorithms used in HR detection, their operational principles and the ongoing research aimed at overcoming the inherent limitations of current technologies. By understanding and addressing these challenges, video-based HR monitoring can become more reliable, paving the way for broader adoption in clinical and non-clinical environments.

3.1.1 | Extract HR Based on ML Advantages and Limitations

Recent advancements in ML have enabled the invention of models that may precisely identify people in VM. VM can be utilised for precise monitoring because these models analyse features in facial ROI and extract data about HR. To improve the accuracy and reliability of HR detection in CHM applications, ML models can recognise facial landmarks, and apply filters to reduce noise in the recovered cardiac signals [96, 100, 101].

A research work by Antink et al. [94] employed a multimodal technique to extract alterations in skin colour and head motions from video footage while concurrently recording a PPG signal using a customised chair. A Bayesian approach was used to aggregate the obtained data to predict HR. Video of the face extracted using the Viola–Jones method and ICA was used by Varsha et al. [83] to demonstrate an alternative method. The FastICA algorithm was employed to isolate the PPG signal from noise, and ML techniques were utilised to combine and enhance the accuracy of the signal for HR measurement. Building on these approaches further, Ghanadian et al. [25] demonstrated an RGB camera-based HR monitoring system that uses ICA first to isolate the biological signal from background noise. Subsequently, a LR model was utilised to choose the channel (R, B or G) that provided the most precise physiological data for estimating HR. In addition, in order to estimate HR, Bashar et al. [102] employed the RF regression algorithm to differentiate between 'noise' and 'non-noise' groups, and the K-means algorithm to remove movement artefacts and noise. Additionally, Elhajjar et al. [103] measured the precision of VM in estimating HR using SVM. VM-based HR was classified as correct or incorrect with an accuracy of 92.1% for an SVM model. The great accuracy of this classification highlights the efficiency of confidence measures in differentiating between accurate and inaccurate HR predictions using VM, making it an essential solution for CHM.

The integration of advanced ML algorithms into CHM systems enhances the accuracy and reliability of HR measurements, making them increasingly valuable for various medical and health applications. As ML and computational techniques continue to evolve, these systems will further improve in precision and robustness, solidifying their role as indispensable tools in modern healthcare [52].

3.1.2 | Extract HR Based on DL Advantages and Limitations

The use of DL to HR extraction from non-invasive measurements has significant potential benefits for medical diagnostics because it allows for real-time data processing and improves accuracy [5]. Although DL models enhance accuracy and enable real-time HR monitoring, their effectiveness can be influenced by factors such as data stability, computational efficiency and privacy considerations [68]. DL's effect on HR monitoring is therefore constantly shifting which has significant ramifications for CHM and patient treatment [12, 17].

In order to go deeper into the topic, Chen et al. [104] extracted the iPPG signal using AMTC. Interbeat intervals (IBI) were used to compute HR after temporal filtering isolated HR frequencies. Improving upon the knowledge gained from their previous study, the author delved deeper into the subject matter [71]. Chen et al. developed DeepPhys which improves HR estimates using attention mechanisms and skin reflex model motion representations. On another approach, Liu et al. [72] used EVM and Gaussian pyramid analysis to increase HR-indicating colour fluctuations in facial footage. Convolution neural networks (CNNs) integrated spatial and temporal data to estimate HR properly. Advanced HR-CNN structures by Spitlik et al. [46] include feature extraction and HR estimation subnetworks. These designs efficiently extract physiological data from facial video frames. Numerous face patch signals were used by Niu et al. [46] to create spatial–temporal maps (ST-maps). These maps were fed into a CNN-RNN regressor. In a related topic, Hu et al. [74] extracted spatial–temporal facial features using 2D and 3D CNN. Then, the aggregation algorithm integrated feature maps into short-segment spatial-–temporal feature maps to extract iPPG signals for accurate HR estimation. Additionally, Zhang et al. [105] estimated HR and RR using the dichroic reflectance model (DRM). To detect physiological signals in tiny fluctuations in reflected light intensity, they used multitasking transient attention networks (MTTS-CAN). Zhao et al. [53] improved iPPG by utilising MSSTNet and multiscale feature (MSF) extraction and spatial and temporal correlation modelling. The Gaussian pyramid was used to separate pulse data from noise, and the layers were inputted into a separable spatio-temporal convolution (SSTC) layer to extract height, width and time features. The dimension separable attention (DSAT) unit assessed these traits and found the heartbeat. Zhao, Hu et al. [106] used CNN with space–time and attention channel technology (ASTNet) to estimate HR from facial films and pulse signals. A study by Zhao et al. [38] introduced the 'multi-aspect ratio sampling' technology. This method creates bounding boxes of various sizes that match people's faces. MSSF was used to combine signals from different scales into the bounding boxes to reduce noise. In addition, camera-PPG HR estimations are also discussed in Ref. [66]. PPG signals must be collected from the right skin ROIs according to the authors. Use PPG features from multiscale picture resolutions to improve measurement. In Ref. [66], the authors constructed GLISNet, a DL model. A local path learns representations in the original scale and a global path learns representations in other sizes to capture multiscale information in this model. For direct training data learning, a hybrid loss was utilised between the two paths. The authors found camera-PPG extraction improvements over the most advanced networks. In real-world low-light circumstances, these gains are very noticeable. Zhao et al. used advanced methods to demonstrate how DL models can

increase real-world accuracy and reliability. Advanced methods including space–time, ASTNet, MSF Extraction, GLISNet and MSSTNet estimated HR from video data. Multimodal data integration is complicated and needs more research to improve its precision and resilience for real-world implementations.

Table 4 shows the results of several studies that used RGB cameras and DL methods to estimate HR. The presence of many methods and advancements in technology is apparent in this field. For example, the MAE and RMSE values are commonly used to evaluate the performance of these systems with smaller values indicating better accuracy. In this context, MAE values range from 0.14 to 22.76 bpm [45, 106, 115], and RMSE values range from 1.81 to 19.15 bpm across the studies [45, 46], suggesting varying levels of precision in HR estimation. Future research should focus on achieving more consistent and accurate results with acceptable error margins in clinical settings. Future research should give priority to the advancement of algorithms capable of real-time operation, validating their performance in dynamic contexts, and addressing practical challenges in the implementation of video-based HR monitoring systems to enhance their effectiveness and usability.

3.2 | Heart Rate Variability (HRV) Algorithms

Building upon the success of ML and DL models in HR estimation, similar strategies were employed for HRV measurement, leveraging spatiotemporal features to improve signal stability under real-world conditions. HRV, which refers to the variation in the time intervals between consecutive heartbeats [116], impacted by a variety of internal and external circumstances for example, nutrition, sleep rate, physical exercise and chronic diseases [117], is a crucial marker of autonomic nervous system function and general physiological resilience [118]. HRV is an advanced biomarker for measuring autonomic reactions to different stimuli and stress [119]. Studies demonstrate how well it measures stress reactions and overall autonomic function [117, 120, 121], demonstrating its use in clinical and wellness contexts by offering crucial insights into personal health dynamics [122]. The importance of HRV in predictive health assessments is supported, for instance, in 2006, by Acharya et al. [120] which shows a significant association between reduced HRV and increased vulnerability to cardiovascular diseases.

An increasing number of mental health facilities are using HRV analysis to better understand the relationship between emotional and psychological stress and HRV alterations [117, 120]. Because decreased parasympathetic activity is often indicated by reduced HRV values, accurate measurements are crucial for recognising and tracking conditions such as anxiety and depression [18, 123]. In sports medicine, HRV is used for checking physical condition and recuperation [124]. HRV changes provide helpful data regarding training readiness and overtraining risk [122]. Furthermore, HRV's importance in the treatment of long-term diseases including diabetes and hypertension has been well-documented. The role of the autonomic system in controlling various disorders can be better understood with the use of HRV [121].

Recent advances in video-based technologies now enable non-invasive HRV monitoring using algorithms that assess subtle physiological changes captured by RGB cameras.

3.2.1 | DL/ML Techniques for HRV Estimation

Following the motivation outlined above, numerous non-contact methods were developed using ML and DL algorithms to extract HRV features from facial videos. The extraction pipeline typically involves three stages [119]: (1) capturing iPPG signals from facial regions, (2) peak detection to identify heartbeat timings and in the third stage, advanced ML algorithms are employed to estimate the normal-to-normal interval (NNI) from the iPPG signal as depicted in Figure 5.

Several ML-based approaches were proposed to enhance the accuracy of NNI estimation and artefact removal. For instance, SVR and bagged decision trees were employed to select noise-free ROIs and predict HRV signals under real-world conditions [18, 125]. Advanced filtering and motion suppression techniques were also integrated to mitigate signal corruption [123]. Additionally, thermal imaging combined with SVR has shown potential in HRV estimation for specific populations [116].

DL-based methods further enhance performance through spatio-temporal learning. PhysNet [38], a 3D-CNN architecture, was designed to extract features across both spatial and temporal dimensions for accurate HRV parameter estimation (LF, HF, LF/HF). ESA-rPPGNet [119] combines attention mechanisms and temporal filtering to improve peak detection. Furthermore, a hybrid PhysNet-RNN model utilising LSTM and ConvLSTM layers has demonstrated improved robustness against motion artefacts. Liu et al. [126] introduced a Dual-GAN framework that achieved promising results for non-contact HRV prediction using adversarial learning. Key studies utilising ML and DL models for HRV estimation are summarised in Table 5, along with their corresponding performance metrics such as MAE, RMSE and Pearson correlation coefficient.

The key challenges in improving the precision of ML/DL-based HRV detection algorithms primarily revolve around several key factors such as peak detection accuracy, noise and artefacts in the signal and dealing with varying lighting conditions [130]. One of the major challenges is accurately detecting the peaks in the rPPG signal, where motion artefacts, lighting disturbances and skin tone variations can lead to distorted or incomplete peak detection, thus compromising HRV accuracy [130]. To address these challenges, advanced peak detection algorithms such as ESA-rPPGNet, which leverage DL for spatial and temporal features, were proposed to improve peak detection accuracy despite noise and artefacts [119]. Additionally, the challenges discussed in Section 2 highlight obstacles related to motion and lighting conditions, where these factors interfere with measurement accuracy [6]. Furthermore, using low-quality cameras or suboptimal setups leads to signal degradation [119]. Solutions to address these challenges include improving noise reduction algorithms and integrating multiple sensor signals to enhance accuracy in changing environments [130].

TABLE 4 | Studies employing RGB cameras using DL algorithm for HR estimation.

Ref.	Year	Algorithm	Camera detail	Graphics card	Dataset	Ground truth	Results
[81]	2018	DeepPhys	Scout scA640 RealSense VF0800 Leap motion	—	MAHNOB-HCI	FlexComp infiniti	MAE: 0.14 SNR: 0.03
[73]	2018	HR-CNN	Logitech HD Webcam C525	—	COHFACE, MAHNOB and PURE	ECG-fitness	RMSE: 19.15 MAE: 14.48 ρ: 0.50
[72]	2019	EVM with CNN	RGB-camera	Core i5-CPU and 8.0 GB RAM	MMSE-HR، MAHNOB-HCI	ECG signal	RMSE: 3.26% SD: −1.68(2.79) ρ: 0.95
[46]	2020	ST-map	Logitech C310 RealSense F200 Huawei P9	Titan 1080Ti GPU	MAHNOB-HCI, MMSE-HR and VIPL-HR	ECG signal	Mean: 0.73 STD: 8.11 MAE: 5.30 RMSE: 8.14 ρ:0.76
[75]	2020	DRM and MTTS-CAN	RGB-camera	—	AFRL and MMSE-HR	Arterial BP monitor and biopac MP150	SNR: 8.64 MAE: 1.45 RMSE: 3.72 ρ: 0.94
[107]	2020	Conv2D and LSTM	Logitech C922	TITAN X GPU	Self-collected	Pulse oximeter (CONTEC CMS60C)	STD: 2.684 MAE: 2.075 RMSE: 2.836
[108]	2022	MSSF	RGB-camera	—	PURE and self-rPPG	Pulse oximeter	SNR: 1.806 MAE: 1.577
[109]	2022	CNN	RGB-camera	—	UBFC-Phys and BIDMC	Pulse oximeter	RMSE: 7.36 MAE: 4.47 ρ: 0.871
[110]	2022	2D-CNN and 3D-CNN	RGB-camera	—	PURE, and COHFACE	—	RMSE: 7.52 MAE: 5.19 ρ: 0.68 PCC: 0.99926
[19]	2022	LSTM	Camera (model D425)	NVIDIA GeForce GTX 1080 Ti	MR-NIRP	Polar H7 HR sensor	RMSE: 3.26
[45]	2023	EfficientPhys	RGB-camera	—	AFRL, UBFC, and MMSE	—	MAE: 1.14 RMSE: 1.81 ρ: 0.99
[106]	2023	ASTNet	Logitech HD Webcam	NVIDIA GTX 1080Ti GPU	PURE, COHFACE and self-rPPG	Pulse oximeter CMS50 E	SNR: 10.19 MAE: 22.76% RMSE:25.43%
[111]	2023	PhysFormer and PhysFormer++	RGB-camera	—	VIPL-HR, MAHNOB-HCI, MMSE-HR and OBF	—	RMSE: 3.88 MAE: 3.23 SD: 3.90
[93]	2023	SSTC	RGB-camera	NVIDIA GeForce GTX 1080Ti	PURE and UBFC-rPPG	Pulse oximeter	MAE: 9.58
[112]	2023	DMD and ST-map	RGB-camera	—	TokyoTech Remote PPG, MR-NIRP and UBFC-rPPG	—	RMSE: 6.37

(Continues)

TABLE 4 | (Continued)

Ref.	Year	Algorithm	Camera detail	Graphics card	Dataset	Ground truth	Results
[113]	2023	STSC and MHFF	RGB-camera	—	UBFC-rPPG, COHFACE and sellfed dataset collected	—	MAE: 1.75
[114]	2023	JAMSNet	RGB-camera	NVIDIA GTX 1080Ti	PURE, UBFC-rPPG and self-rPPG	—	SNR: 10.65 MAE: 0.63 RMSE: 1.07 MAPE: 0.87
[66]	2024	GLISNet	RGB-camera	NVIDIA GTX 1080Ti	PURE, UBFC-rPPG	—	SNR: 8.627 MAE: 1.050 RMSE: 1.422

Abbreviations: MAPE, Mean Absolute Percentage Error; MHFF, Multi-Hierarchical Feature Fusion Module; PCC, Pearson correlation coefficient; SD, standard deviation; SNR, signal-to-noise ratio; STSC, Spatio-Temporal Stack Convolution Module.

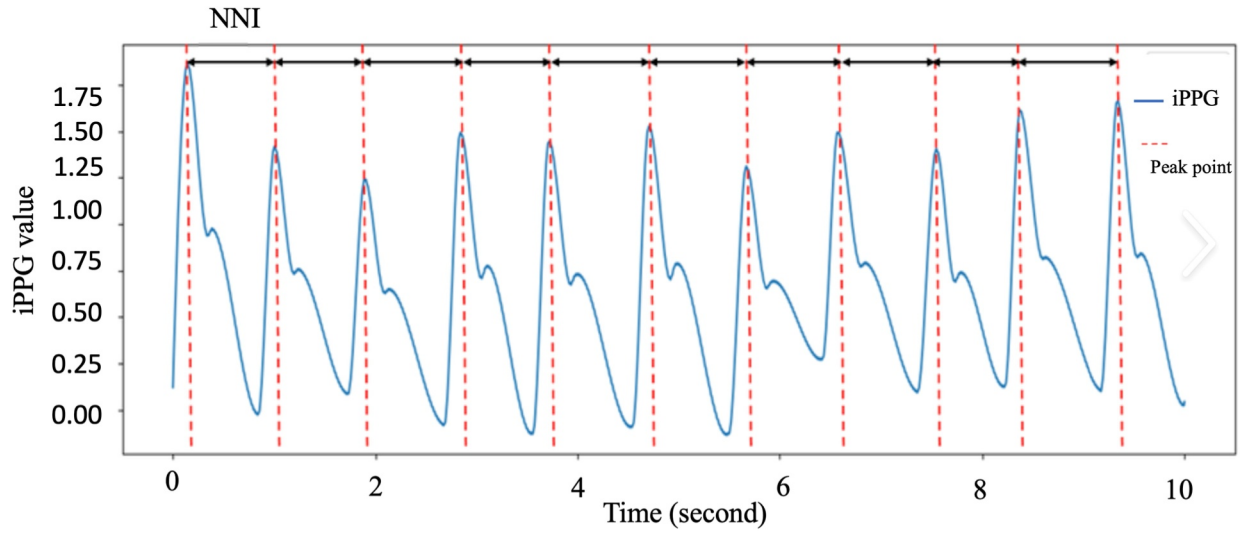

FIGURE 5 | Displays the normal-to-normal interval obtained from the iPPG signal [119].

3.3 | Respiratory Rate (RR) Algorithms

RR, which is the fundamental measure of how often a person breathes, is commonly quantified as the number of breaths per minute [131]. Depending on the age, normative rates might vary and usually fall between 12 and 20 breaths per minute [132, 133]. This indication is significant as it provides insight into an individual's respiratory and metabolic well-being [36]. The fluctuations in RR can serve as a reliable and responsive signal of distress and probable hypoxia as well as a severe sickness. Variations from these averages can indicate possible health problems: lower values could indicate brain damage or neurological impairments, whereas higher values could indicate respiratory distress, certain infections or metabolic acidosis [134, 135]. Continuous monitoring of RR is essential for enhancing the delivery of drugs or therapies [27].

In clinical environments such as critical care, anaesthesia and postoperative settings, accurate RR assessment is essential for timely intervention and improved patient outcomes [27, 134, 136]. However, traditional methods like manual counting and capnography often fall short for continuous monitoring because of limitations in accuracy and practicality [137]. This has led to growing interest in non-invasive, algorithm-driven monitoring systems that provide real-time data [138, 139]. Physical activities such as walking or talking can introduce motion artefacts, making it difficult to capture reliable RR signals. To address this, video-based systems increasingly rely on ML/DL algorithms that enhance robustness against noise and movement as discussed in the following sections.

Recent advances in ML-based techniques have significantly enhanced RR estimation by addressing motion artefacts, enhancing signal processing and integrating multi-source data fusion. RR estimation using iPPG and RGB-based video monitoring is improved through ML algorithms that enable robust motion tracking, adaptive filtering and multimodal data fusion. Estimating RR from subtle facial and chest movements presents challenges because of external noise, voluntary body motion and environmental lighting variations [140]. 3D-CNN architectures extract spatiotemporal features from video sequences, capturing fine-grained respiratory patterns by analysing pixel-wise intensity changes across multiple frames [56, 141]. Additionally, multimodal ML approaches combining RGB-based motion analysis with iPPG-derived respiratory signals improve measurement robustness, ensuring consistent RR estimations in varying conditions, including occlusions and low-light environments [142]. LSTM-based filtering techniques refine

TABLE 5 | Displays a number of studies that have employed DL methods to predict HRV using iPPG data.

Ref.	Year	Algorithm	Camera detail	Dataset	Ground truth	Results
[127]	2013	3D CNN	RGB camera 720*576 at 25 fps	KTH data	—	Precision: 0.78 recall: 0.0340
[38]	2019	PhysNet-3DCNN and RNN based PhysNet (LSTM and ConvLSTM)	RGB videos recorded at 60 fps with a resolution of 1920 × 2080	OBF and MAHNOB-HCI	ECG signal	ACC: 80.22% SP: 81.71%
[47]	2020	MST-map	RGB-camera	VIPL-HR, OBF, and MMSE-HR	BVP signals	—
[128]	2020	3D CNN	RGB-camera	MAHNOB-HCI	ECG signal	—
[129]	2021	PulseGAN	RGB-camera	UBFC-rPPG, PURE, and MAHNOB-HCI	Pulse oximeter	AVNN: 41.19, 40.45, 41.63% SDNN: 37.53, 44.29, 58.41%
[126]	2021	DuAl-GaN	RGB-camera	UBFC-rPPG, VIPL-HR, PURE	CMS50E	*r*: 0.891 for LF. *r*: 0.881 for LF/HF
[119]	2022	ESA-rPPGNet	RGB-camera	UBFC-rPPG and PURE	CMS50E oximeter	For LF and HF: RMSE = 0.0546, $r = 0.9018$ LF/HF: RMSE: 0.2191, *r*: 0.9424

Abbreviations: ACC, accuracy; AVNN, average of all normal-to-normal intervals; SDNN, standard deviation of all NN intervals; SP, specificity.

respiratory signal detection by distinguishing between voluntary motion artefacts and true breathing patterns, reducing false estimations and enhancing overall accuracy [140]. These advancements align with the challenges outlined in Section 2 (challenges in video-based vital signs monitoring), particularly in addressing motion artefacts, real-time processing constraints and the impact of environmental variability on RR detection. Suriani et al. in [141] and Qayyum et al. in [140] employed CNNs, Z. Liu and et al. [56] utilised LSTM and 1D-CNN within 3D-CNN, and W. Chen [143] employed DeepPhys to predict the RR. The utilisation of these algorithms was effective in addressing the difficulties associated with obtaining respiratory signals, particularly in detecting chest movements and subtle face changes caused by breathing. These cutting-edge techniques are being developed to surpass the constraints of current methods and adjust to evolving patient monitoring situations, signifying a significant breakthrough in the field of non-contact respiratory monitoring.

3.3.1 | Extract RR Based on ML Advantages and Limitations

VM technique retrieves RR measurements from the nose and thoracic/abdominal regions [56, 144]. Nevertheless, studies indicate that the use of a thermal camera is necessary in order to accurately assess nose temperature [145]. The camera must maintain an uninterrupted and unobstructed line of sight to the nostrils which restricts the possible camera angles and positions [88, 88, 146, 147]. A particular breathing rhythm results from temperature variations during inhaling and exhaling [148]. Thermal cameras are too expensive for families and have poor image quality [32, 136]. Some authors estimated RR using RGB camera face motions [39, 40, 142, 149, 150]. The second area has been extensively studied. This method requires watching the chest move while breathing. The thoracic wall moves during breathing and exhalation to expand and contract the lungs. The rib cage and abdomen of the chest wall move separately during breathing. Inhalation shifts the rib cage 3–5 mm forward and up, rib elevation is helped by 1–2 mm lateral motions [32]. The abdomen moves ventrally, making it circular during inhalation. The rib cage has a lower dorsoventral-transverse diameter ratio than the abdomen during expiration. Circularity is greater in the belly than in the rib cage [151]. Figure 6 shows VM for estimating RR on RGB-camera, thermal camera, or combined RGB-camera and thermal camera using ML, DL or combined ML & DL techniques.

Recently, VM techniques for remote RR estimation, particularly in contactless methods, have shown promise [39]. RGB cameras record facial and upper torso motions, making these methods a convenient and non-intrusive alternative to traditional measurement equipment [152]. Firstly, Schrumpf, Monch et al. [153] calculated colour channel average pixel values to generate face motion signals. Instead of tracking, they employed chest movement to preserve consistency between facial and chest motion inputs. Using iPPG and mobility data, empirical mode decomposition (EMD) determined respiratory modulation. A weighted median with the signal-to-noise ratio was used to obtain the RR. Additionally, Gwak, Vatanparvar, Kuang et al. [154] estimated RR using LR using facial and upper body video frames to account for movement-induced distortions. Since lung motion impacts head skin motion, Pediaditis et al. [155] utilised

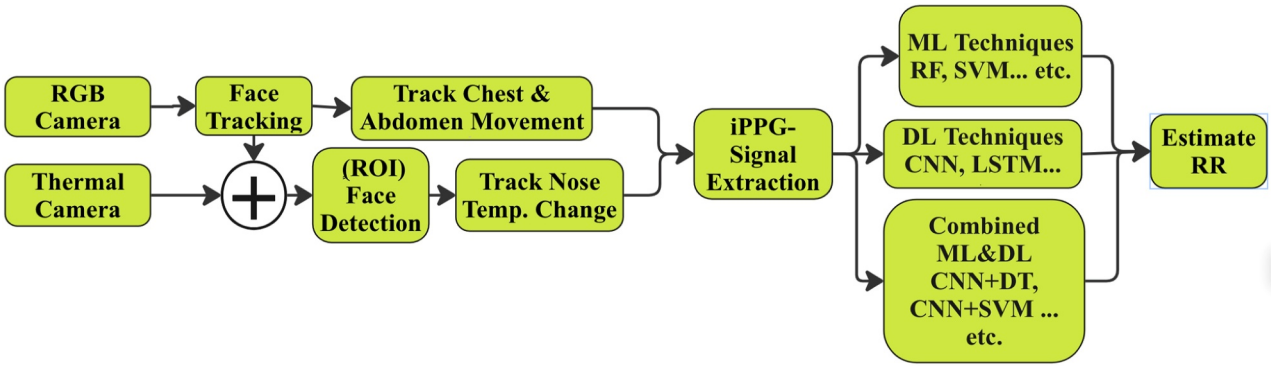

FIGURE 6 | Diagram showing how thermal/colour cameras with ML/DL can predict RR.

an RGB camera with head motions alone to estimate RR. They trained the RF classifier on head movement features to detect apnoea. Expanding on these fundamental principles, a study [152] used video plethysmography (VPPG), created by the Kang Lee Lab at the University of Toronto, to capture haemoglobin variations under the epidermis utilising advanced optical signal processing and ML. VPPG, like iPPG, assessed RR using iPad face video without patient interaction. Expanding on this approach, Gwak, Vatanparvar, Zhu et al. [142] introduced a multi-modal technique that blends face motions with RGB camera-derived iPPG data from skin colour fluctuations. This method used a binary DT algorithm to calculate the RR from the respiratory signal. Building on this foundation, Ghodratigohar et al. [156] extracted the iPPG signal using CEEMDAN and RF algorithms. They also used k-star and rotation forest algorithms to determine the best classifier for RR estimate performance. Similarly, Gwak, Vatanparvar, Zhu et al. [149] used an RGB camera to record head motions and use ML algorithms to extract signal features to diagnose RR and apnoea.

3.3.2 | Extract RR Based on DL Advantages and Limitations

By employing DL algorithms and an RGB camera, researchers have devised multiple methods to enhance the precision of contactless rate measurements [86]. This advancement has resulted in wider applications in CHM and clinical diagnostics [101, 115, 157]. Researchers have made substantial progress in improving the precision and dependability of non-contact methods for estimating RR from video data, based on various studies. Initially, Zhan et al. [158] emphasised that analysing pixel motions, as opposed to pixel intensity changes, resulted in more precise estimations of respiratory activity. This suggests a transition towards dynamic analytic methodologies. Building upon these discoveries, Suriani et al. [141] employed facial movies for fitness purposes, utilising a combination of face detection and iPPG approaches in conjunction with the CNNs algorithm. Despite a small margin of error, their method represented a significant advancement in the integration of DL algorithms with physiological monitoring. In another study [159], the accuracy of contactless respiratory rate estimates was significantly increased by using enhanced thermal imaging and super-resolution DL networks. More specialised methods were introduced by M. Chen, Zhu et al. [40] and Gastel et al. [37], who used Eulerian video magnification (EVM) to amplify chest movements and the tasks-constrained deep convolutional network (TCDCN). By employing optical flow techniques, these models not only allowed for more precise evaluation but additionally enhanced the level of accuracy of the iPPG that was returned.

Leveraging these innovations, Z. Liu and et al. [56] developed the respiratory signal extractor model using data from ICU patients. This model utilises a combination of LSTM, dense and 1D-CNN inside a 3D-CNN architecture to accurately forecast RR and extract breathing waveforms. This model showcased the abilities of spatiotemporal and unsupervised learning techniques while also pushing the boundaries of accuracy and dependability in RR monitoring. Similarly, Ren et al. [150] and Kolosov et al. [36] investigated the concurrent prediction of HR and RR utilising sophisticated DL approaches with the iPPG signal. The authors of their study emphasised the advantages of both precision and hardware efficiency, subsequently highlighting the wide range of uses and adaptations of these technologies.

On top of that, applying ML methods for classification such as DT and SVM classifiers, along with Bayesian-optimised artificial neural networks (AHNs), in the approaches suggested by Jagadev et al. [160] and Brieva, Ponce et al. [161], has not only enhanced the precision of estimating RR from video data but has also made substantial advancements towards the overarching objective of analysing extensive clinical data using non-contact RR estimation algorithms. These studies collectively demonstrate substantial and gains advancements in the development of health monitoring systems that are simultaneously more accessible and more accurate.

Table 6 provides an overview of key studies that have explored the use of camera-based technology for RR monitoring, highlighting the diversity of methodologies, sensor types and performance metrics. These studies leverage various DL architectures, including CNN and recurrent models, applied to RGB, infrared and thermal imaging data. The evaluation primarily relies on metrics such as MAE, RMSE and correlation coefficients which offer insight into estimation accuracy. Based on reported results, MAE values range from 0.14 [162] to 14.48 bpm [46, 73], whereas RMSE values vary between 0.066 [56] and 19.15 bpm [47, 108]. R spans from 0.45 [161] to 0.995 [166], reflecting significant variability in model performance. Sensitivity values range from 60.7% [164] to 90% [163], whereas specificity is reported at a maximum

TABLE 6 | Summary of the articles considering camera-based technology for RR monitoring.

Ref.	Year	Method	Camera detail	Dataset	Ground truth	Results
[162]	2018	CNN	Camera VF0800 780 × 580, 61 FPS Near-IR 640 × 240	RGB Video I, RGB Video II, MAHNOB-HCI, and infrared video	—	MAE: 0.14 SNR: 10.8
[143]	2019	DeepPhys	Web-camera	Self-collected	—	RMSE: 2.1
[163]	2019	Multi-task DL	JAI 3-CCD AT-200CL	Self-collected	Philips patient monitor	MAE: 3.5 Accurate: 90%
[40]	2019	TCDCN	Canon camera	Self-collected	PPG sensor, respiratory belt	M: 5.92% SD: 0.04 ± 2.18 RMSE: 2.16
[140]	2019	CNN	Logitech C920 HD pro webcam	Two different datasets	Wrist Ox2 model 3150 wrist-worn	MAE: 3.5
[159]	2019	DRCN	Thermal cameras Flir SC3000	Self-collected	—	RMSE: improves by 0.15 bpm for 8-bit and by 0.84 bpm for 16-bit MAE: Reduced by 0.63 bpm
[164]	2020	SQI, PSD, ACR, CPSD and MUSIC	Dual camera DFK23U618, FLIR A315	Self-collected	Respiration effort belt	*r*: 0.87 Sensitivity: 60.7% Specificity: 86.4%
[158]	2020	CNN	RGB CMOS camera USB M2ST036H	HNU respiration	—	—
[165]	2021	EVM and pyramid decomposition	Point grey Flea 3 GigE	Self-collected	Polysomnography	Error reduced from 15.7% to 1.5%
[166]	2022	3D-CNN	RGB camera	PURE, UBFC-rPPG, OBF, MR-NIRP and MMSE-HR	ECG	RMSE: 1.00 MAE: 0.64 *r*: 0.995
[155]	2022	3D-CNN	IR camera	Self-collected	Respironics alice sleep diagnostics	MAE: 2.29
[56]	2023	3D-CNN and LSTM	RGB camera	Self-collected	BROADSIMS TeleTouch	RMSE: 0.066 *r*: 0.501
[167]	2023	EVM	Canon	COHFACE	—	MAE: 1.62 PCC: 0.45
[161]	2023	CNN and AHN	Canon EOS 1300D	Self-collected	—	Percentage error: 2.19 ± 2.1

of 86.4% [164]. For SpO2 estimation, percentage errors have been observed between 0.066 [56] and 2.19 ± 2.1 [161]. These values highlight the wide range of performance across different models, emphasising the need for standardised benchmarks to ensure consistency and clinical applicability.

However, the lack of standardised benchmarks and publicly available datasets presents challenges in directly comparing model performance. Future advancements should focus on developing standardised evaluation frameworks to enable meaningful comparisons, facilitating more reliable and clinically relevant applications of contactless RR estimation.

3.4 | Blood Pressure (BP) Estimation

Blood pressure (BP) is a critical physiological marker for cardiovascular health, typically measured in millimetres of mercury (mmHg) [168]. Clinically, hypertension is diagnosed when systolic BP (SBP) exceeds 140 mmHg or diastolic BP (DBP) exceeds 90 mmHg [169]. Traditional measurement methods, including invasive catheterisation and non-invasive cuff-based sphygmomanometers, although accurate, are impractical for continuous and contactless monitoring [170]. Recent advancements in computer vision and ML/DL techniques have enabled non-invasive BP estimation from facial video-based iPPG

signals, offering promising alternatives for remote healthcare applications [11, 61].

These approaches leverage facial or palmar regions to extract physiological features from iPPG signals, enabling the estimation of systolic and diastolic BP without physical contact. As illustrated in Figure 3b, fine-grained fluctuations in skin tone are analysed to infer cardiovascular dynamics. These methods enhance estimation accuracy and reliability, supporting their potential use in continuous monitoring and personalised health interventions.

3.4.1 | ML Approaches for BP Estimation

Recent studies have increasingly adopted ML-based approaches for contactless blood pressure (BP) estimation using camera-derived iPPG signals. For example, Zhou et al. [171] extracted peak and valley features from iPPG waveforms and combined them with BMI to estimate BP via linear regression. Building on this, Viejo et al. [172] analysed raw video data (RVA) to detect colour fluctuations across RGB channels, developing two ML models one for forecasting time-series data and another for estimating HR and BP highlighting a link between vital signs and behavioural patterns.

In another approach, a study by [173] utilised smartphone-based transdermal optical imaging (TOI) to extract facial iPPG signals, applying advanced ML techniques to estimate SBP and DBP in a large population. Similarly, Yawle et al. [163] implemented a multiple regression model on PPG signals captured by a smartphone camera to predict BP values. Rong et al. [11] further expanded this line of research by extracting 26 iPPG features and training four ML models, where SVR yielded the best performance with standard deviations of 3.35 (SBP) and 2.58 mmHg (DBP), and MAE values of 9.97 and 7.59 mmHg, respectively.

Extending the scope, Chen et al. [165] employed facial and hand videos with ensemble learning algorithms to estimate BP, proposing a framework for future non-invasive monitoring systems. Collectively, these efforts underscore the growing potential of ML in remote BP estimation while also emphasising the need for improved generalisability and robustness to support clinical translation.

3.4.2 | DL-Based BP Estimation

Recent research has leveraged DL models to extract blood pressure (BP) from iPPG signals by learning complex waveform features. For instance, Rong et al. [73] integrated area, slope and energy characteristics into a neural network to estimate BP, achieving strong performance. Similarly, other studies employed LSTM-based architectures to infer arterial stiffness-related features from facial video signals [169].

To enhance model generalisation, the multi-site physiological monitoring (MSPM) dataset [160] provides synchronised camera recordings from multiple body locations, enabling more robust estimation of BP, RR and pulse transit time (PTT). Table 7 presents a comparative overview of DL models used for BP estimation, reporting MAE values ranging from 4.43 mmHg (DBP, U-Net) to 7.05 mmHg (SBP, ImageNet pre-trained models) [49, 51, 170].

Lin et al. [176] proposed a dual-path DL system combining video magnification and CNNs to extract HR, PTT and waveform features from facial ROIs. B. F. Wu, Chiu et al. [180] applied InfoGAN and an encoder–decoder architecture to generate synthetic iPPG data from facial regions, enhancing model training using the MIMIC III dataset. These methods highlight the potential of DL in remote BP monitoring yet their generalisation remains constrained by environmental variability and data scarcity. As illustrated in Figure 7, their approach localised two primary ROIs on the upper and lower face based on nasal landmarks to extract iPPG signals for BP prediction.

To overcome the limited availability of training data, several strategies have been adopted including the generation of synthetic signals via GANs [180], the use of transfer learning [51] and the integration of multimodal inputs (e.g., ECG and thermal imaging) [166]. Expanding diverse annotated datasets remains essential for improving model robustness and clinical relevance.

3.5 | Blood Oxygen Saturation (SpO2) Estimation

Blood oxygen saturation (SpO2) serves as a vital indicator of pulmonary and circulatory efficiency in CHM systems. Accurate SpO2 assessment is essential in both acute and chronic care contexts, prompting the development of advanced non-contact methods. SpO2 is a crucial physiological indicator used to monitor and treat chronic respiratory and cardiovascular conditions [181]. Equation (1) states that the oxygenation status of an individual can be determined by comparing their total haemoglobin (HP) to their oxygenated haemoglobin (HbO2) [182]. Pulse oximetry is the most reliable method for measuring the level of oxygen saturation in the blood known as SpO2 [183]. Traditionally, it consists of wavelengths within the red and infrared range as depicted in Figure 2. The distinction between (HbO2) and (Hb) in the red and blue wavelengths indicates that these colour channels provide valuable data for predicting SpO2 using optophysiological principles [184].

Continuous monitoring of SpO2 is necessary, especially in critical care units, as low SpO2 readings can indicate potentially fatal conditions such as heart failure or chronic obstructive pulmonary disease [185]. Typically, SpO2 levels within the range of 95%–100% are regarded as normal [186]. On the other hand, hypoxia is a medical condition characterised by insufficient oxygen flow to the body, which can be identified when oxygen levels fall below 95% [187, 188]. Non-invasive methods of detecting SpO2 levels have two advantages: preventing contamination and avoiding issues with blood circulation [186]. By monitoring patients around the clock, medical professionals can promptly adjust therapies to achieve optimal oxygen levels, improving patient outcomes and lowering the likelihood of complications [189]. This practice enables healthcare professionals to promptly determine appropriate courses of action

TABLE 7 | Synopsis of camera-based technology for BP monitoring.

Ref.	Year	Method	Camera detail	Dataset	Results
[174]	2018	ANN	RGB camera	Self-collected	Error rate: For the afternoon session: SBP: 9.62%, DBP: 11.63%. For the evening session: SBP: 8.4%, DBP: 11.18%.
[175]	2020	MLP	Webcam with 30 fps	—	RMSE: (SBP: 7.942, DBP: 7.912) mmHg. MAE: (SBP: 6.556, DBP: 6.372) mmHg.
[11]	2021	SVR	Webcam	Self-collected	STD: (SBP: 3.35, DBP: 9.97) mmHg. MAE: (SBP: 2.58, DBP: 7.59) mmHg.
[158]	2021	AlexNet, ResNet and LSTM	USB camera (IDS UI-3040CP, 32 fps)	MIMIC-B	—
[176]	2022	InfoGAN and ENC-DEC	—	Self-collected	MAE for (SBP: 9.13 and DBP: 8.76) mmHg.
[177]	2022	CNN modified from ResNet	RGB cameras	Self-collected	MAE for (SBP: 11.54, DBP: 8.09) mmHg.
[178]	2022	CNN based on ResNet and CBAM	RGB camera (DFK33UX174)	—	PCC: 0.85 and MAE: 5.4 mmHg.
[179]	2023	U-Net	RGB cameras	MIMIC-II, rPPG	MAE for (DBP: 4.43, SBP: 6.9) mmHg.
[159]	2023	RNN	—	MIMIC	ME for SBP: 0.92 with SD: 7.79. ME for DBP: 0.68 SD: 4.94.
[170]	2023	CNN	—	—	RMSE: For SBP: 4.11, DBP: 3.75. MAPE: For SBP: 3.73, DBP: 4.95.
[167]	2024	Transfer DL using ImageNet	—	MIMIC-I, MINIC-II and MIMIC-III	STD for SBP: 8.027882 and DBP: 6.013052. MAE for SBP: 7.052803 and DBP: 5.616028.
[171]	2024	ENC-DEC	Logitech C920	MIMIC III	MAE for SBP: 8.78 mmHg, DBP: 6.16 mmHg.

Abbreviation: ME, Minimum error.

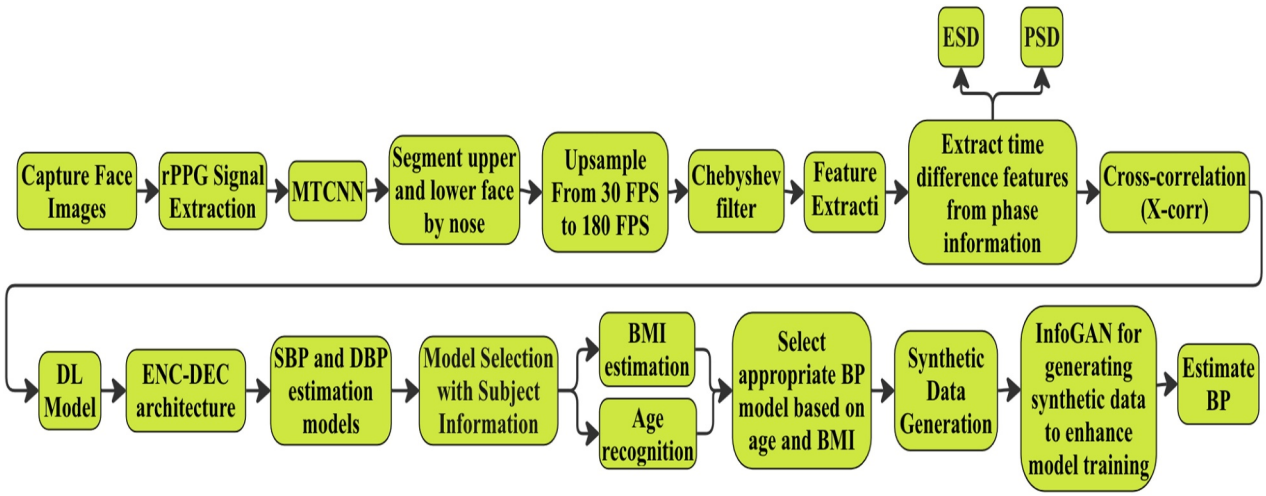

FIGURE 7 | Illustrates the overall structure employed to measure BP by examining two separate ROIs on the face.

and interventions, ensuring sufficient oxygen delivery and averting potential complications [190].

Recent advancements have introduced the use of RGB cameras to estimate SpO2 using facial or hand videos based on ML/DL algorithms have been proposed, enriching the field with tools that leverage video data at multiple wavelengths. A distance of approximately 1 m is typically employed in research [191–193] with a framerate of 30 frames per second. Some studies have used lower framerates such as 10 or 25 fps [182, 191, 194]. RGB digital cameras are employed to ascertain SpO2 values, often with a video resolution of 640 × 480 pixels per frame [195, 196].

ROI is obtained from different areas of the face such as the forehead, left cheek, right cheek and lips [35, 193, 196]. The absorbance ratio is determined by measuring the intensity of light at two specific wavelengths (660 and 940 nm) using the red and blue colour channels [30, 192]. The SpO2 value can be estimated using the ratio of ratios (RR) value. The RR is calculated using the following relation:

$$RR = \frac{AC\lambda 1/DC\lambda 1}{AC\lambda 2/DC\lambda 2} = \frac{R\lambda}{R\lambda} = \frac{ACred/DCred}{ACblue/DCblue} = \frac{Rred}{Rblue} \quad (1)$$

The variable AC represents the pulsatile component divided by the non-pulsatile component (DC) at the wavelengths λ_1 and λ_2 as shown in Figure 8. RR refers to the ratio of the absorbance at the two wavelengths [196] as shown in Figure 5. The SpO2 value can be approximated based on the RR value, as there is a close linear correlation between both as stated in the following equation [182, 194]:

$$SpO2 = RR \cdot \alpha + \beta \quad (2)$$

where the values of a and b are found by analysing the LR for each participant. Specifically, the variable 'a' represents the gradient or slope of the estimated regression line, whereas 'b' represents the *y*-intercept which is the point where the regression line contacts the *y*-axis [194].

The field of SpO2 estimation has been enhanced by recent advancements in RGB cameras which utilise video-based tools and leverage data from different wavelengths. Groundbreaking research has utilised various advanced ML/DL models, including SVM [194], CNN [184], MultiPhys [86], ST-map [198] and CL-SpO2Net [199], to effectively analyse video data obtained from facial [200] and hand [186, 201] regions. These advanced algorithms play a crucial role in greatly improving the precision of SpO2 measurements, thus transforming the way vital signs are monitored without invasive methods and pushing the limits of patient care in healthcare technology.

3.5.1 | Extract SpO2 Based on ML Advantages and Limitations

ML algorithms have attracted considerable attention due to their potential in non-invasive medical monitoring, namely in the estimation of SpO2 levels. KNN algorithm has been employed with RGB cameras [202], whereas SVM has been utilised with hand footage captured by smartphones, showcasing the adaptability of ML applications [203]. As shown in another study [204], videos of people's faces shot using cellphones in natural light have been analysed using the SVM algorithm. This showed common ways to use SpO2 data. The techniques for estimating SpO2 can be classified into feature-based, colour signal sequence-based and video-based methods, each employing distinct model inputs. Additionally, Hamoud et al. [205] use a special method that involves using a pre-trained CNN to look at face video. The features that were retrieved are then used to teach an XGBoost regressor ML programme what to do. Results from two datasets, VIPL-HR and UBFC-rPPG, show how well the approach performs and how accurate it is in contactless SpO2 estimation. When comparing UBFC-rPPG and VIPL-HR the MAE was 0.84 and 1.17, respectively.

3.5.2 | Extract SpO2 Based on DL Advantages and Limitations

Deep learning (DL) algorithms have significantly advanced non-invasive SpO_2 monitoring. Gastel et al. [137] used CNNs alongside POS and CHROM algorithms to minimise motion artefacts and estimate SpO_2 based on light absorption ratios. Similarly, EVM-based techniques [196] achieved results comparable to standard pulse oximeters. Other studies [186] applied 3D-CNNs to video frames of the palm and hand, whereas Stogiannopoulos et al. [206] employed 3D-CNN and ViViT on infrared facial videos, demonstrating strong performance across varying sample sizes.

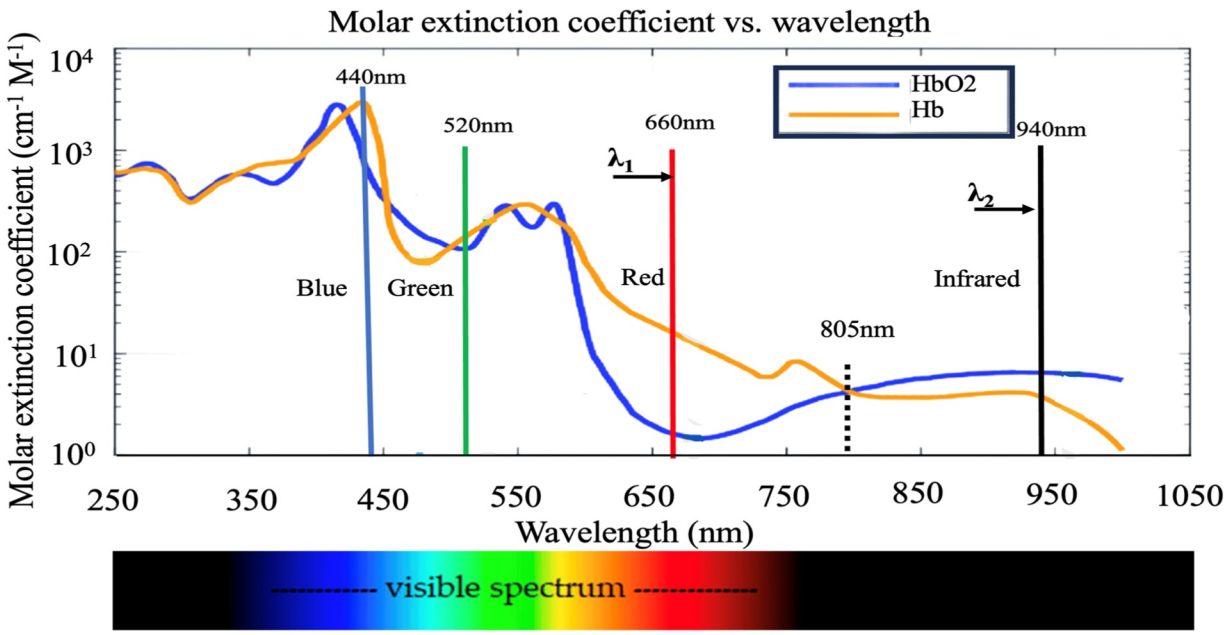

FIGURE 8 | shows the extinction coefficient curves of haemoglobin, illustrating its absorption abilities at different wavelengths [184, 197].

Akamatsu et al. [198] integrated EVM with LS-SVM, RBF and DNNs to enhance estimation, and later introduced synthetic data generation with DL to improve CNN performance for HR and SpO_2 [86]. Hu, Wu et al. [206] proposed the RCA network for facial video analysis, whereas Cheng et al. [206] utilised ST-maps with 2D-CNNs on the VIPL-HR dataset. Peng et al. [199] developed CL-SpO2Net, combining ML and DL for small datasets. Sun et al. [199] introduced CCSpO2Net, using temporal and spatial feature extraction modules, to lower MAE.

Jalaja and Kavitha [95] designed a hybrid deep ensemble reinforcement learning model (HDERL) for early anomaly detection. Collectively, these efforts highlight the diverse and effective use of DL in SpO_2 estimation. Althoug many models report MAE below 1.2%, their clinical applicability still warrants further investigation. Table 8 presents a comparative overview of these studies.

ML–DL is being used in a new way for presentation at CHM. This approach leverages DL and ML algorithms to analyse changes in skin tone observed by RGB cameras, correlating them with oxygen saturation levels [31, 189]. A key advantage of this method is its ability to continuously and remotely monitor patients in various environments, such as homes and hospitals, thereby enhancing comfort and improving adherence to medication for chronic heart and lung conditions [189]. In addition to these advantages, the current research on SpO_2 remains limited compared to HR analysis as depicted in Figure 9. Existing DL models must be trained on multiple datasets to ensure strong generalisation across different demographics and environmental conditions [209]. Additionally, researchers need

TABLE 8 | Summary of the articles considering camera-based technology for SpO2 monitoring.

Ref.	Year	Method	Camera detail	Dataset	Ground truth	Results
[187]	2019	CNN	iPhone 6s	Self-collected	Pulse oximeter	MAR: 2.02%
[140]	2019	CNN	Logitech C920 HD	Two different datasets	Pulse oximeter	SpO2: 98.33.
[109]	2022	MultiPhys	RGB camera	UBFC-phys and BIDMC	Pulse oximeter	MAE: 0.551 RMSE: 0.879 ρ: 0.450
[207]	2022	MTCNN	RGB camera	PURE and VIPL-HR	Pulox CMS50 E and CONTEC CMS60 C	MAEs < 2%
[201]	2022	Linear multiple regression	Logicool C922n	Self-collected from hand	Fingertip photoelectric pulse	ME: 2.46% Max. Correlation coefficient: 0.495
[28]	2022	End-to end CNN	eco 274cvge	PURE	Pulse oximeter (pulox cms50e)	MAE: 1.54 RMSE: 1.66
[189]	2023	CNN	Smartphone cameras	Self-collected from hand	—	ρ: 0.46 RMSE: 0.36 IQR: 2.32
[208]	2023	ST- map	Logitech C920n	Self-collected	Pulse oximeter	CNN-Filter, MAE: 1.17, RMSE: 1.40, CorrCoef: 0.474 CNN-EndToEnd, MAE: 1.18, RMSE: 1.40, CorrCoef: 0.496
[31]	2023	ViViT with 3D-CNN	Google Nest Cam	Self-collected	JPD-500D ControlBios Oxicore pulse oximeter.	RMSE: 1.331 R^2 score: 0.465
[193]	2023	3D-CNN	RGB camera	Self-collected	Pulse oximeter	MAE: 1.123% RMSE: 1.632 MSE: 2.662
[206]	2024	(ST-Map) 2D - CNNs	RGB camera	VIPL-HR	PolarOH1	MAE: 1.274% RMSE: 1.71%
[199]	2024	CL-SpO2Net	Logitech C920 HD	Forensics++, UBFC and MVPD	CMS50 E pulse oximeter.	MAE: 0.85% in the stable environment. MAE: 1.13% with lighting fluctuations. MAE: 1.20% in the facial rotation situation.
[95]	2024	HDERL	RGB camera	TokyoTechrPPG, PURE and COHFACE	—	MAE: 1.176% for the PURE dataset

Abbreviations: IQR, interquartile range; MSE, mean square error.

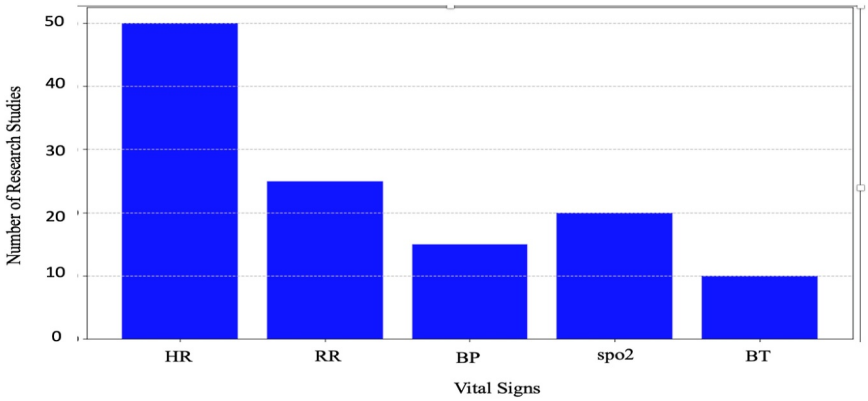

FIGURE 9 | Number of articles considering VM with ML/DL for vital signs parameters (2016—July 2024).

to develop more sophisticated algorithms to address uncertainties and mitigate issues related to label and annotation inconsistencies [210]. Another critical factor is the lack of SpO_2 estimation for specific locations in clinical and physiological contexts when individuals are in motion. Therefore, implementing SpO_2 estimation in real-time remains a crucial area for future development [30, 207].

3.6 | Body Temperature (BT) Estimation

Monitoring body temperature (BT) is an integral component of contactless health monitoring systems, offering real-time insights into systemic inflammatory or metabolic states. Recent advancements allow estimation of BT through visual and infrared signals. In medicine, BT is a vital sign that reflects the body's ability to generate and dissipate heat, maintaining a stable internal environment [63]. Nevertheless, BT of an individual is subject to variation and influenced by factors such as gender, age, time of day, level of physical activity, health condition (including illness and menstrual cycle in females), the specific location on the body where the measurement is obtained, the individual's level of consciousness (awake, asleep or sedated) and emotional state [211]. The normal BT for a healthy adult typically ranges from 36.1°C (97°F) to 37.2°C (99°F) [212]. BT above this range is considered high and may indicate fever or hyperthermia, often associated with infections, inflammations and other medical conditions. Conversely, a BT below 36.1°C (97°F) is considered low and can signal hypothermia which may result from prolonged exposure to cold, metabolic disorders or certain medical conditions [144].

Precise assessment of BT is crucial, particularly in medical environments, as it enables prompt detection of illnesses and monitoring of their advancement. Presently [213], the majority of touch technologies employed for the purpose of measuring BT, such as medical thermometers and thermistors, need substantial human exertion, time and resources. This not only imposes a heavy workload on healthcare professionals but also squanders valuable resources. Furthermore, in order to guarantee precision in the measurement, the individual's activities will be limited, and the procedure may result in discomfort [213]. Non-contact BT measurement employs thermal cameras [63, 145], RGB cameras [211, 214] and IR cameras [116, 215] in contrast to conventional touch approaches. In recent times, sophisticated technologies have been integrated with ML–DL algorithms, such as KNN [216], CNN-SVM [211] and logistic regression [217], employing very precise and non-intrusive methods to evaluate thermal data and forecast BT. This offers a practical substitute for conventional thermometers.

However, extracting BT using camera and DL technologies presents several limitations and challenges. From a hardware perspective, high-quality IR cameras can be expensive and may require regular calibration to maintain accuracy. Environmental factors such as ambient temperature, humidity and the presence of reflective surfaces can also affect the accuracy of thermal readings. Algorithmically, DL models require large datasets for training which must be diverse and representative to ensure reliability across different populations and conditions. Additionally, developing these models involves addressing issues such as overfitting and ensuring that the algorithms can generalise well to new unseen data. Real-time processing and integration into existing healthcare systems also pose significant challenges, requiring robust software solutions and substantial computational resources. Despite these challenges, integrating camera and ML–DL technologies holds significant potential for improving the accuracy and convenience of BT measurement, ultimately enhancing healthcare outcomes [157, 218].

Temperature reading of the body entails multiple phases. Typically, the temperature of the human body is measured without physical contact by scanning the entire face or specific frontal features such as the eye, nose, forehead and cheek. ML or DL models autonomously identify the target area and convert its coordinates into ROI, as shown in Figure 10a,b, either by taking the overall ROI temperature or the peak temperature of the forehead [29, 41, 157, 164].

Multiple research investigations are conducted by writers on the topic of assessing BT utilising ML approaches. Multiple research investigations are conducted by writers on the topic of assessing BT utilising Ml approaches. To differentiate between healthy and infected individuals, Dagdanpurev et al. [216] utilised KNN algorithm. Additionally, K. Yao et al. [214] propose a KNN algorithm that utilise the Raspberry Pi Camera Module v2. The camera module, which is equipped with a high-quality 8-megapixel Sony IMX219 sensor, is utilised to estimate BT. However, they first estimated BT using an RGB camera for face

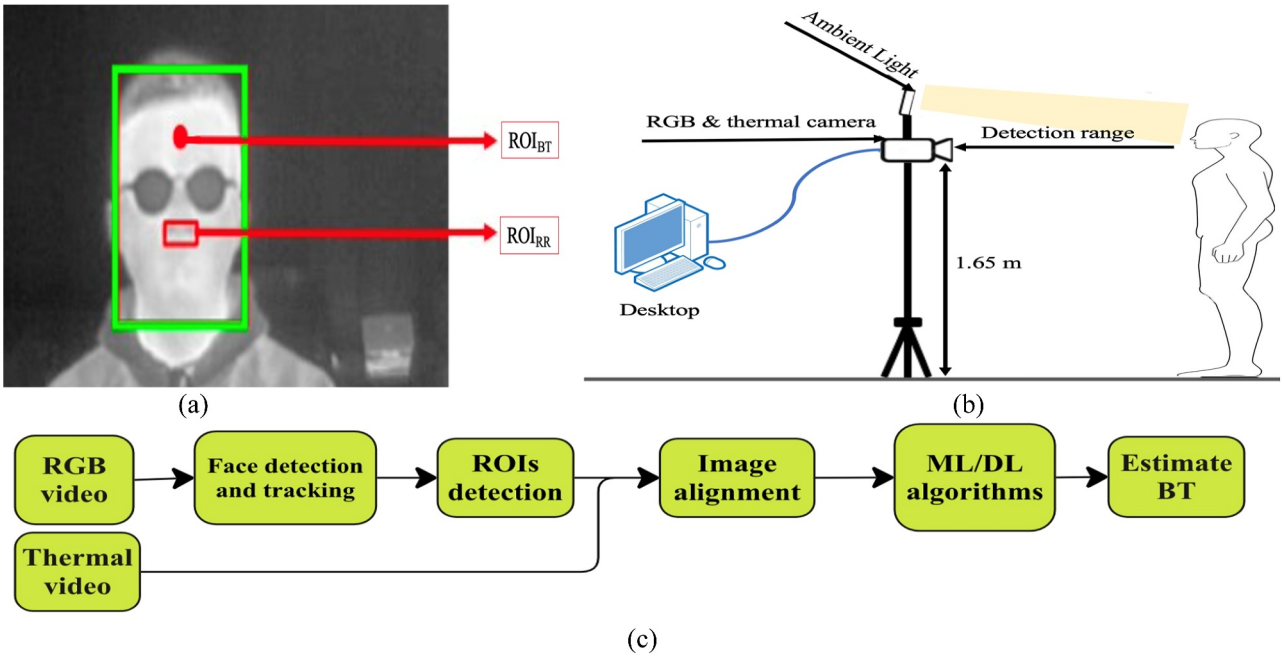

FIGURE 10 | (a) Illustration of vital signs measurement experiment setup [88], (b) locations of ROI (RR) and ROI (BT) on simultaneous thermal image [88] and (c) workflow for BT estimation using RGB camera and thermal cameras.

detection and then switched to a thermal camera due to the latter's poor clarity for face detection. A study of interest [219] investigated the use of video-based temperature (V-TEMP) technology to detect elevated skin temperature using regular video footage. This approach utilised the angular pattern of skin reflectance which was discovered to be analogous to the BT derived from facial skin. Researchers classified skin according to its optical characteristics using a threshold classifier for both low and high temperatures. Furthermore, Unursaikhan et al. [217] monitored RR and HR using a VISC-camera while utilising a thermal camera to monitor BT. After, logistic regression analysis was used to tell the difference between people who had Coronavirus based on their vital signs all within a 10-s window. As an option Lazri et al. [145] used infrared thermographs (IRTs), which are also called thermal cameras, to measure high BT and RR. BT can be reliably detected in areas close to the inner corner of the eyes, and RR may be reliably detected in nasal regions, according to research using the single-shot multibox detector (SSD) approach for facial recognition. In addition, Jakkaew, and Onoye [220] used a thermal camera to analyse thermal pictures to determine respiration. To find ROI, the researchers found the BT that was the highest. They then used a method based on the absolute difference between frames to find out if there were any significant body movements during sleep. In another study, Y. Zheng et al. [221] have proposed the development of mobile applications that employ CNNs to extract face characteristics and apply multiclass-SVM to accurately assess BT using facial videos.

These studies demonstrate the progress made in non-contact temperature and vital sign monitoring based on ML, illustrating the effectiveness of integrating thermal imaging with ML and DL approaches to provide precise health evaluations. The consistent focus on particular facial areas for accurate measurements, the incorporation of logistic regression for the swift identification of infections and the inventive application of video-based temperature detection highlight the various methodologies and their potential to enhance public health surveillance.

To the topic of quantifying BT on DL methods, the authors moreover carried out a few investigations. For instance, one such study, conducted by J. Wei J.-W. Lin et al. [63] presented an automated continuous body temperature measurement (CBTM) system utilising a FLIR thermal camera. Using SSD with MobileNet, they achieved real-time extraction of the subject's forehead temperature. The results show that the MAE and RMSE are 0.375°C and 0.439°C, respectively. This demonstrates the ability of advanced thermal imaging methods combined with ML models to increase the accuracy and efficacy of continuous BT monitoring.

Similarly, C. Lyra, Mayer et al. [215] employed YOLOv4-Tiny algorithms to identify the head and chest regions in IRT frames. They applied an optical flow algorithm to extract RR from the chest region and used LR statistical methods to estimate trends in BT over time from a patient's head. Their work highlights the need to use many physiological parameter measures to offer thorough CHM.

Figure 11 below, that show outlines the challenges associated with ML/DL methods including hardware and software challenges. This comprehensive view underscores the evolution and complexities of BT measurement in clinical settings.

4 | Conclusions

This article presents a literature review of previously published studies on the use of VM with ML and DL algorithms for the CHM of an individual's vital signs. The second section on VM was organised into six sections, each titled after a vital sign: HR,

HRV etc. For each vital sign, we conducted thorough research on the techniques involved. The initial segment for each sign employs ML, whereas the subsequent segment for the same sign encompasses DL-based methodologies, followed by an elucidation of the advantages, open problems and limitations associated with each biomarker.

The research conducts a comparative analysis of studies that prioritise vision-based approaches, taking into account several criteria such as the learning algorithm utilised, camera kinds and specifications, the dataset employed and the accuracy of monitoring vital signs, assessed through various criteria such as MAE, RMSE and STD. We examine the difficulties and limitations of each approach, concluding each section on vital signs by addressing the issues encountered in obtaining these essential indicators.

In this research review, we noticed a scarcity of studies using novel dataset methodologies for measuring and estimating SpO2 and BP. Typically, this research relies on pre-existing dataset methods which are only applicable in ideal environments and practically tested in clinical settings. Furthermore, we assert that both ML and DL methodologies necessitate abundant training data, advanced hardware specifications and prolonged processing durations to achieve desired outcomes. These obstacles impede the models from attaining maximum efficiency. Although the testing environment was optimal and did not accurately represent real-world conditions, the outcomes in the field of vital sign estimation were encouraging. Continuous unobtrusive monitoring enhances the quality of daily healthcare.

5 | Future Research Directions

Integrating ML/DL in video-based CHM offers a transformative approach to non-invasive health assessment. It is particularly advantageous for real-time and remote medical evaluations. As shown in Figure 12 below, we explore potential research

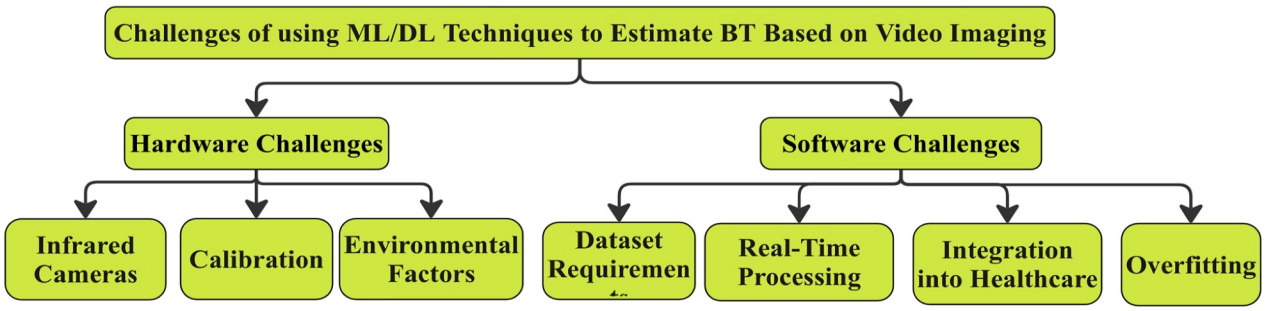

FIGURE 11 | Challenges in leveraging ML/DL techniques for estimating BT through video imaging.

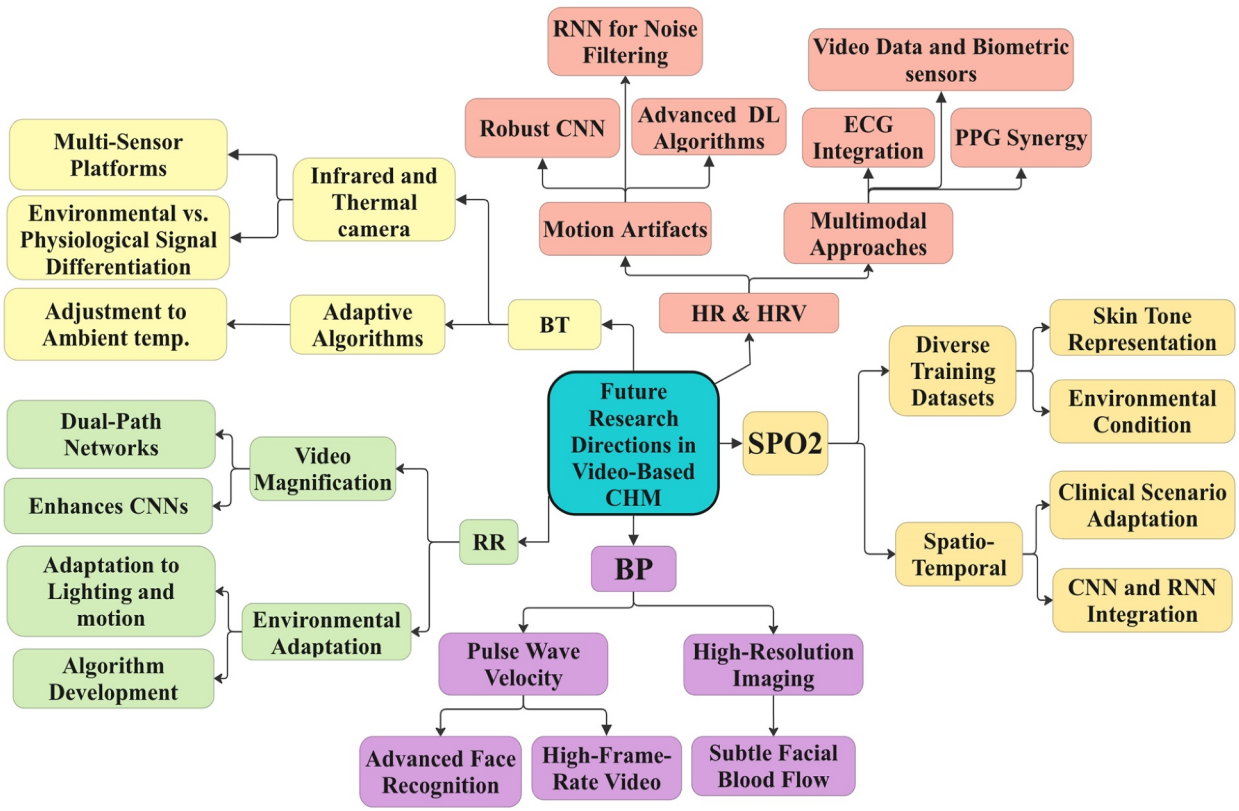

FIGURE 12 | Diagram of future directions for vital signs monitoring in video-based CHM using ML/DL.

avenues to refine the estimation accuracy of vital signs, each accompanied by novel and specific enhancements.

5.1 | HR and HRV

It is important to improve sophisticated DL algorithms that increase the precision of data during movement in order to address the problem of motion artefacts in HR monitoring. Future studies should focus on enhancing these algorithms to maximise the accuracy of signal fidelity and effectively identify heartbeats, even when there is mobility. By combining video footage with biometric sensor data, such as ECG or PPG, we can investigate the potential of obtaining more accurate assessments of HR and HRV. Improved PPG Approaches: These methods make better use of iPPG by utilising high-resolution video recording to accurately identify and assess minute variations in skin tone that correspond to heartbeats. By integrating DL techniques, this methodology can improve the precision of remote HR readings. This will improve comprehension of cardiovascular stability and stress reactions.

5.2 | RR

Future research in RR estimation should prioritise the improvement of video magnification approaches by integrating sophisticated CNN and dual-path networks. This would enhance the procedure of extracting and verifying the precision of iPPG signals from facial videos with varying levels of motion velocities. Adjusting algorithms is crucial for properly managing environmental elements including lighting conditions and background movements.

5.3 | SpO2

Improving the accuracy of non-contact measurements, the measurement of SpO2 depends on the creation of varied training datasets that include a wide range of skin tones and environmental conditions. Combining CNNs with RNN architectures could enhance the precision of forecasting spatial and temporal changes, hence enhancing the reliability of SpO2 readings in different demographic and clinical scenarios. Developing and implementing calibration techniques for non-contact SpO2 readings that account for differences in skin tone and ambient lighting conditions in order to improve accuracy. The use of SpO2 data with other vital signs in telehealth applications can enhance diagnostic precision and patient care. To address the scarcity of datasets for SpO2 and BP estimation, future research should focus on developing more diverse datasets through multispectral imaging and generating synthetic data using GANs [180]. Additionally, adopting a multimodal learning approach that integrates RGB, thermal, PPG and ECG signals can enhance estimation accuracy. Moreover, establishing a labelled video dataset based on real SpO2 and BP signals, while ensuring demographic diversity, will improve DL models' ability to extract relevant features and enhance prediction accuracy. Developing and implementing calibration techniques for non-contact SpO2 readings that account for differences in skin tone and ambient lighting conditions in order to improve accuracy.

5.4 | BP

The analysis of pulse wave velocities obtained from high-frame-rate video data can significantly improve the accuracy of non-invasive BP monitoring. By combining these measurements with advanced facial recognition technologies, it is feasible to attain more precise and impartial assessments of BP. The model is trained consistently to accurately evaluate the effectiveness of detecting BP from the facial area as demonstrated by B. F. Wu, Chiu et al. [180].

5.5 | BT

Advanced methodologies are used in the development of CHM systems to improve the precision of non-contact BT measurement. Using an RGB camera to collect iPPG signals from changes in facial blood flow over time provides a novel method for detecting BT. This strategy relies on the association between BT and blood circulation. When the temperature increases, the body increases blood flow to the outer layers of the skin to release heat. Conversely, when the temperature decreases, the body slows blood flow to conserve heat. ML and DL models can be created and trained to establish a connection between iPPG features and body temperature, allowing for the prediction of temperature using a thermometer. Subsequently, classification algorithms can be utilised to ascertain if the temperature falls within the defined range of typical values or exceeds it. Furthermore, progress in IR and thermal imaging technology should prioritise distinguishing between temperature signals related to the environment and those related to the body's physiology. To do this, it is necessary to create advanced algorithms that can separate the heat pattern of the human body from the surrounding temperature effects. These algorithms may be integrated into multi-sensor platforms to improve accuracy.

The advancement of ML/DL in CHM has been hindered by the challenges associated with putting theoretical models into reality. When applied in the real world, unlike in a lab, unexpected problems are common. Accordingly, situational models are crucial for their practical efficacy. With these advancements, CHM may be able to outfit a patient's home with a network of cameras to keep tabs on their vitals in a more streamlined and thorough manner. Based on an analysis of current research's emerging trends and challenges, these future paths aim to build a more robust and precise framework for non-invasive health monitoring.

Author Contributions

Alaa Hajr: formal analysis, methodology, writing – original draft. **Bahram Tarvirdizadeh:** conceptualization, methodology, project administration. **Khalil Alipour:** conceptualization, methodology. **Mohammad Ghamari:** supervision.

Acknowledgements

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

References

1. Y. Jiang, F. Jiang, Y. Zhi, et al., "Artificial Intelligence in Healthcare: Past, Present and Future," *Stroke and Vascular Neurology* 2, no. 4 (2017): 230–243, https://doi.org/10.1136/svn-2017-000101.

2. S. Doraiswamy, A. Abraham, R. Mamtani, and S. Cheema, "Use of Telehealth During the COVID-19 Pandemic: Scoping Review," *Journal of Medical Internet Research* 22, no. 12 (December 2020): e24087, https://doi.org/10.2196/24087.

3. P. H. King, "Contactless Vital Signs Monitoring," *IEEE Pulse* 13, no. 6 (2022): 39–40, https://doi.org/10.1109/MPULS.2022.3227812.

4. S. Zaunseder, A. Trumpp, D. Wedekind, and H. Malberg, "Cardiovascular Assessment by Imaging Photoplethysmography – a Review," *Biomedical Engineering/Biomedizinische Technik* 63, no. 5 (2018): 617–634, https://doi.org/10.1515/bmt-2017-0119.

5. R. J. Lee, S. Sivakumar, and K. H. Lim, "Review on Remote Heart Rate Measurements Using Photoplethysmography," *Multimedia Tools and Applications* 83, no. 15 (2024): 44699–44728, https://doi.org/10.1007/s11042-023-16794-9.

6. A. Gudi, M. Bittner, and J. van Gemert, "Real-Time Webcam Heart-Rate and Variability Estimation With Clean Ground Truth for Evaluation," *Applied Sciences* 10, no. 23 (2020): 8630, https://doi.org/10.3390/app10238630.

7. N. Molinaro, E. Schena, S. Silvestri, et al., "Contactless Vital Signs Monitoring From Videos Recorded With Digital Cameras: An Overview," *Frontiers in Physiology* 13 (February 2022), https://doi.org/10.3389/fphys.2022.801709.

8. K. P. Abrisham, K. Alipour, B. Tarvirdizadeh, and M. Ghamari, "Predicting Vascular Age Using PPG Signals and Machine Learning Algorithms: A Non-Invasive Approach for Early Cardiovascular Risk Detection," in *2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM)*, (2023), 72–77, https://doi.org/10.1109/ICRoM60803.2023.10412455.

9. H. Mohammadi, B. Tarvirdizadeh, K. Alipour, and M. Ghamari, "Blood Pressure Estimation Through Photoplethysmography and Machine Learning Models," in *2023 9th International Conference on Control, Instrumentation and Automation (ICCIA)*, (2023), 1–6, https://doi.org/10.1109/ICCIA61416.2023.10506396.

10. A. Rostami, B. Tarvirdizadeh, K. Alipour, and M. Ghamari, "Real-Time Stress Detection From Raw Noisy PPG Signals Using LSTM Model Leveraging TinyML," *Arabian Journal for Science and Engineering* 50, no. 10 (2024): 6959–6981, https://doi.org/10.1007/s13369-024-09095-2.

11. K. Motaman, K. Alipour, B. Tarvirdizadeh, and M. Ghamari, "A Stress Detection Model Based on LSTM Network Using Solely Raw PPG Signals," in *2022 10th RSI International Conference on Robotics and Mechatronics (ICRoM)*, (2022), 485–490, https://doi.org/10.1109/ICRoM57054.2022.10025256.

12. Y. Hasanpoor, K. Motaman, B. Tarvirdizadeh, K. Alipour, and M. Ghamari, "Stress Detection Using PPG Signal and Combined Deep CNN-MLP Network," in *2022 29th National and 7th International Iranian Conference on Biomedical Engineering (ICBME)*, (2022), 223–228, https://doi.org/10.1109/ICBME57741.2022.10052957.

13. J. Brieva, E. Moya-Albor, H. Ponce, and A. Escobedo-Gordillo, "Contactless Video-Based Vital-Sign Measurement Methods: A Data-Driven Review," in *Data-Driven Innovation for Intelligent Technology: Perspectives and Applications in ICT*, eds. H. Ponce, J. Brieva, O. Lozada-Flores, L. Martínez-Villaseñor, and E. Moya-Albor (Springer Nature Switzerland, 2024), 1–38, https://doi.org/10.1007/978-3-031-54277-0_1.

14. Y. Sun and N. Thakor, "Photoplethysmography Revisited: From Contact to Noncontact, From Point to Imaging," *IEEE Transactions on Biomedical Engineering* 63, no. 3 (2016): 463–477, https://doi.org/10.1109/TBME.2015.2476337.

15. R. H. Y. So, P. K. Man, K. L. Cheung, et al., "Blood Pressure Measurement: From Cuff-Based to Contactless Monitoring," *Healthcare* 10, no. 10 (2022): 1–42, https://doi.org/10.3390/healthcare10102113.

16. W. Wang, A. C. Den Brinker, S. Stuijk, and G. De Haan, "Algorithmic Principles of Remote PPG," *IEEE Transactions on Biomedical Engineering* 64, no. 7 (July 2017): 1479–1491, https://doi.org/10.1109/TBME.2016.2609282.

17. A. Ni, A. Azarang, and N. Kehtarnavaz, "A Review of Deep Learning-Based Contactless Heart Rate Measurement Methods," *Sensors* 21, no. 11 (June 2021): 3719, https://doi.org/10.3390/s21113719.

18. X. Yu, T. Laurentius, C. Bollheimer, S. Leonhardt, S. Member, and C. H. Antink, "Noncontact Monitoring of Heart Rate and Heart Rate Variability in Geriatric Patients Using Photoplethysmography Imaging," *IEEE Journal of Biomedical and Health Informatics* XX, no. XX (2020): 1–12, https://doi.org/10.1109/JBHI.2020.3018394.

19. J. Przybyło, "A Deep Learning Approach for Remote Heart Rate Estimation," *Biomedical Signal Processing and Control* 74 (2022): 103457, https://doi.org/10.1016/j.bspc.2021.103457.

20. D. Shao, C. Liu, and F. Tsow, "Noncontact Physiological Measurement Using a Camera: A Technical Review and Future Directions," *ACS Sensors* 6, no. 2 (February 2021): 321–334, https://doi.org/10.1021/acssensors.0c02042.

21. D. F. Swinehart, "The Beer-Lambert Law," *Journal of Chemical Education* 39, no. 7 (1962): 333–335, https://doi.org/10.1021/ed039p333.

22. M. Rong and K. Li, "A Blood Pressure Prediction Method Based on Imaging Photoplethysmography in Combination With Machine Learning," *Biomedical Signal Processing and Control* 64 (2021): 102328, https://doi.org/10.1016/j.bspc.2020.102328.

23. L. Saikevičius, V. Raudonis, G. Dervinis, and V. Baranauskas, "Non-Contact Vision-Based Techniques of Vital Sign Monitoring: Systematic Review," *Sensors* 24, no. 12 (2024): 3963: [Online], https://www.mdpi.com/1424-8220/24/12/3963.

24. T. P. Pagano, V. R. Santos, Y. da S. Bonfim, et al., "Machine Learning Models and Videos of Facial Regions for Estimating Heart Rate: A Review on Patents, Datasets, and Literature," *Electronics* 11, no. 9 (2022): 1473, https://doi.org/10.3390/electronics11091473.

25. H. Ghanadian, M. Ghodratigohar, and H. Al Osman, "A Machine Learning Method to Improve Non-Contact Heart Rate Monitoring Using an RGB Camera," *IEEE Access* 6, no. October (2018): 57085–57094, https://doi.org/10.1109/ACCESS.2018.2872756.

26. D. McDuff, "Camera Measurement of Physiological Vital Signs," *ACM Computing Surveys* 55, no. 9 (2023): 1–40, https://doi.org/10.1145/3558518.

27. K. Vatanparvar, M. Gwak, L. Zhu, J. Kuang, and A. Gao, "Respiration Rate Estimation From Remote PPG via Camera in Presence of Non-Voluntary Artifacts," in *2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, (2022), 1–4, https://doi.org/10.1109/BSN56160.2022.9928485.

28. J. Lee and C. Park, "Remote SpO2 Estimation Using End-to-End CNN Model," in *2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, (2022), 1–3, https://doi.org/10.1109/ICCE-Asia57006.2022.9954776.

29. Y. T. Chuang, S. J. Lai, T. F. Chang, and Y.-H. Lin, "Image-Based Skin Temperature and Pulse Rate Measuring System," in *2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)* (IEEE, 2020), 1–2, https://doi.org/10.1109/ICCE-Taiwan49838.2020.9258206.

30. R. C. De Fatima Galvao Rosa and A. Betini, "Noncontact SpO2 Measurement Using Eulerian Video Magnification," *IEEE Transactions on Instrumentation and Measurement* 69, no. 5 (2020): 2120–2130, https://doi.org/10.1109/TIM.2019.2920183.

31. T. Stogiannopoulos, G. A. Cheimariotis, and N. Mitianoudis, "A Study of Machine Learning Regression Techniques for Non-Contact SpO2 Estimation From Infrared Motion-Magnified Facial Video," *Information* 14, no. 6 (2023): 301, https://doi.org/10.3390/info14060301.

32. C. Massaroni, D. S. Lopes, D. Lo Presti, E. Schena, and S. Silvestri, "Contactless Monitoring of Breathing Patterns and Respiratory Rate at the Pit of the Neck: A Single Camera Approach," *Journal of Sensors* 2018 (2018): 1–13, https://doi.org/10.1155/2018/4567213.

33. L. Xi, W. Chen, C. Zhao, X. Wu, and J. Wang, "Image Enhancement for Remote Photoplethysmography in a Low-Light Environment," in *Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020*, (2020), 761–764, https://doi.org/10.1109/FG47880.2020.00076.

34. L. Xi, X. Wu, W. Chen, J. Wang, and C. Zhao, "Weighted Combination and Singular Spectrum Analysis Based Remote Photoplethysmography Pulse Extraction in Low-Light Environments," *Medical Engineering & Physics* 105, (2022): 103822, https://doi.org/10.1016/j.medengphy.2022.103822.

35. S. Chen, K. L. Wong, T. T. Chan, Y. Wang, R. H. Y. So, and J. W. Chin, "An Image Enhancement Based Method for Improving rPPG Extraction Under Low-Light Illumination," *Biomedical Signal Processing and Control* 100, (2025): 106963, https://doi.org/10.1016/j.bspc.2024.106963.

36. D. Kolosov, V. Kelefouras, P. Kourtessis, and I. Mporas, "Contactless Camera-Based Heart Rate and Respiratory Rate Monitoring Using AI on Hardware," *Sensors* 23, no. 9 (2023): 4550, https://doi.org/10.3390/s23094550.

37. M. Van Gastel, S. Stuijk, S. Overeem, J. P. Van Dijk, M. M. Van Gilst, and G. De Haan, "Camera-Based Vital Signs Monitoring During Sleep – A Proof of Concept Study," *IEEE Journal of Biomedical and Health Informatics* 25, no. 5 (2021): 1409–1418, https://doi.org/10.1109/jbhi.2020.3045859.

38. S. Networks and Z. Yu, "Remote Photoplethysmograph Signal Measurement From Facial Videos Using Spatio-Temporal Networks," *arXiv:1905.02419v2* (2019).

39. M.-A. Fiedler, P. Werner, M. Rapczyński, and A. Al-Hamadi, "Deep Face Segmentation for Improved Heart and Respiratory Rate Estimation From Videos," *Journal of Ambient Intelligence and Humanized Computing* 14, no. 7 (2023): 9383–9402, https://doi.org/10.1007/s12652-023-04607-8.

40. M. Chen, Q. Zhu, H. Zhang, M. Wu, and Q. Wang, "Respiratory Rate Estimation From Face Videos," in *2019 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2019 - Proceedings*, (2019), 1–4, https://doi.org/10.1109/BHI.2019.8834499.

41. S. Coşar, Z. Yan, F. Zhao, T. Lambrou, S. Yue, and N. Bellotto, "Thermal Camera Based Physiological Monitoring With an Assistive Robot," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2018), 5010–5013, https://doi.org/10.1109/EMBC.2018.8513201.

42. Y. Deshpande, S. Thapa, A. Sarkar, and A. L. Abbott, "Camera-Based Recovery of Cardiovascular Signals From Unconstrained Face Videos Using an Attention Network," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, (June 2023), 5975–5984, https://doi.org/10.1109/CVPRW59228.2023.00636.

43. Food and Drug Administration. Pulse Oximeters - Premarket Notification Submissions [510(k)s] Guidance for Industry and Food and Drug Administration Staff, Vol. 510, (2015): [Online], 1–19, https://www.fda.gov/RegulatoryInformation/Guidances/ucm341718.htm.

44. A. Fawzy, H. Ali, P. H. Dziedzic, et al., "Skin Pigmentation and Pulse Oximeter Accuracy in the Intensive Care Unit: A Pilot Prospective Study," *American Journal of Respiratory and Critical Care Medicine* 210, no. 3 (August 2024): 355–358, https://doi.org/10.1164/rccm.202401-0036LE.

45. X. Liu, B. Hill, Z. Jiang, S. Patel, and D. McDuff, "EfficientPhys: Enabling Simple, Fast and Accurate Camera-Based Cardiac Measurement," in *Proceedings - 2023 IEEE Winter Conference on Applications of Computer Vision, WACV 2023*, (2023), 4997–5006, https://doi.org/10.1109/WACV56688.2023.00498.

46. R. Spetlik, V. Franc, J. Cech, and J. Matas, "Visual Heart Rate Estimation With Convolutional Neural Network," in *British Machine Vision Conference 2018* (BMVC, August 2018), https://dblp.uni-trier.de/db/conf/bmvc/bmvc2018.html#SpetlikFCM18.

47. Y. Qiu, Y. Liu, J. Arteaga-Falconi, H. Dong, and A. El Saddik, "EVM-CNN: Real-Time Contactless Heart Rate Estimation From Facial Video," *Ieee Transactions on Multimedia* 21, no. 7 (2019): 1778–1787, https://doi.org/10.1109/TMM.2018.2883866.

48. B.-F. Wu, B.-J. Wu, B.-R. Tsai, and C.-P. Hsu, "A Facial-Image-Based Blood Pressure Measurement System Without Calibration," *IEEE Transactions on Instrumentation and Measurement* 71 (2022): 1–13, https://doi.org/10.1109/TIM.2022.3165827.

49. E. C. L. Hyeji Lim, J. Baek, J. Baek, and E. C. Lee, "RGB Camera-Based Blood Pressure Measurement Using U-Net Basic Generative Model," *Electronics* 12, no. 3771 (2023): 3771, https://doi.org/10.3390/electronics12183771.

50. M. Chan, L. Zhu, K. Vatanparvar, M. Gwak, J. Kuang, and A. Gao, "Estimating SpO2 With Deep Oxygen Desaturations From Facial Video Under Various Lighting Conditions: A Feasibility Study," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (EMBC, 2023), 1–5, https://doi.org/10.1109/EMBC40787.2023.10340025.

51. O. A. Omer, M. Salah, L. Hassan, A. Abdelreheem, and A. M. Hassan, "Video-Based Beat-By-Beat Blood Pressure Monitoring via Transfer Deep-Learning," *Applied Intelligence* 54, no. 6 (2024): 4564–4584, https://doi.org/10.1007/s10489-024-05354-9.

52. S. S. B. A. Rahman, T. Debnath, D. Kundu, et al., "Machine Learning and Deep Learning-Based Approach in Smart Healthcare: Recent Advances, Applications, Challenges and Opportunities," *American Institute of Mathematical Sciences* 11, no. 1 (2024): 58–109, https://doi.org/10.3934/publichealth.2024004.

53. M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *IEEE Transactions on Biomedical Engineering* 58, no. 1 (2011): 7–11, https://doi.org/10.1109/TBME.2010.2086456.

54. W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote Plethysmographic Imaging Using Ambient Light," *Optics Express* 16, no. 26 (2008): 21434–21445: [Online], https://api.semanticscholar.org/CorpusID:7782171.

55. G. de Haan and V. Jeanne, "Robust Pulse Rate From Chrominance-Based rPPG," *IEEE Transactions on Biomedical Engineering* 60, no. 10 (2013): 2878–2886, https://doi.org/10.1109/TBME.2013.2266196.

56. Z. Liu, B. Huang, C.-L. Lin, et al., "Contactless Respiratory Rate Monitoring for ICU Patients Based on Unsupervised Learning," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, (2023), 6005–6014, https://doi.org/10.1109/CVPRW59228.2023.00639.

57. D. Botina-Monsalve, Y. Benezeth, and J. Miteran, "Performance Analysis of Remote Photoplethysmography Deep Filtering Using Long

Short-Term Memory Neural Network," *BioMedical Engineering Online* 21, no. 1 (2022): 1–28, https://doi.org/10.1186/s12938-022-01037-z.

58. D. J. Chen and W.. McDuff, "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks," *ArXiv* abs/1805.0 (2018): 356–373, https://doi.org/10.1007/978-3-030-01216-8_22.

59. Z. Yu, X. Li, and G. Zhao, "Remote Photoplethysmograph Signal Measurement From Facial Videos Using Spatio-Temporal Networks," *arXiv:1905.02419* (2019), https://arxiv.org/abs/1905.02419.

60. Z. Yu, W. Peng, X. Li, X. Hong, and G. Zhao, "Remote Heart Rate Measurement From Highly Compressed Facial Videos: An End-To-End Deep Learning Solution With Video Enhancement," in *Proceedings of the IEEE International Conference on Computer Vision*, (2019), 151–160, https://doi.org/10.1109/ICCV.2019.00024.

61. F. Schrumpf, P. Frenzel, C. Aust, G. Osterhoff, and M. Fuchs, "Assessment of Deep Learning Based Blood Pressure Prediction From PPG and RPPG Signals," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, (2021), 3815–3825, https://doi.org/10.1109/CVPRW53098.2021.00423.

62. B. Lin, J. Tao, J. Xu, L. He, N. Liu, and X. Zhang, "Estimation of Vital Signs From Facial Videos via Video Magnification and Deep Learning," *iScience* 26, no. 10 (2023): 107845, https://doi.org/10.1016/j.isci.2023.107845.

63. J.-W. Lin, M.-H. Lu, and Y.-H. Lin, "A Thermal Camera Based Continuous Body Temperature Measurement System," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, (2019), 1681–1687, https://doi.org/10.1109/ICCVW.2019.00208.

64. A. V Moço, S. Stuijk, and G. de Haan, "New Insights Into the Origin of Remote PPG Signals in Visible Light and Infrared," *Scientific Reports* 8, no. 1 (2018): 8501, https://doi.org/10.1038/s41598-018-26068-2.

65. S. He, Z. Han, C. Iglesias, V. Mehta, and M. Bolic, "A Real-Time Respiration Monitoring and Classification System Using a Depth Camera and Radars," *Frontiers in Physiology* 13, no. March (2022): 1–14, https://doi.org/10.3389/fphys.2022.799621.

66. C. Zhao, M. Zhou, Z. Zhao, B. Huang, and B. Rao, "Learning Spatio-Temporal Pulse Representation With Global-Local Interaction and Supervision for Remote Prediction of Heart Rate," *IEEE Journal of Biomedical and Health Informatics* 28, no. 2 (2024): 609–620, https://doi.org/10.1109/JBHI.2023.3252091.

67. Y. Benezeth, D. Krishnamoorthy, D. J. Botina Monsalve, K. Nakamura, R. Gomez, and J. Mitéran, "Video-Based Heart Rate Estimation From Challenging Scenarios Using Synthetic Video Generation," *Biomedical Signal Processing and Control* 96 (2024): 106598, https://doi.org/10.1016/j.bspc.2024.106598.

68. I. Odinaev, K. L. Wong, J. W. Chin, R. Goyal, T. T. Chan, and R. H. Y. So, "Robust Heart Rate Variability Measurement From Facial Videos," *Bioengineering* 10, no. 7 (July 2023): 851, https://doi.org/10.3390/bioengineering10070851.

69. J. Tang, X. Li, J. Liu, X. Zhang, Z. Wang, and Y. Wang, "Camera-Based Remote Physiology Sensing for Hundreds of Subjects Across Skin Tones," *arXiv Preprint arXiv:2404.05003* (2024).

70. X. Tang, J. Chen, K. Wang, et al., "MMPD: Multi-Domain Mobile Video Physiology Dataset," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBS, 2023), https://doi.org/10.1109/EMBC40787.2023.10340857.

71. R. Stricker, S. Muller, and H. M. Gross, "Non-Contact Video-Based Pulse Rate Measurement on a Mobile Service Robot," in *IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions*, (2014), 1056–1062, https://doi.org/10.1109/ROMAN.2014.6926392.

72. G. Heusch, A. Anjos, and S. Marcel, "A Reproducible Study on Remote Heart Rate Measurement," (2017) [Online], http://arxiv.org/abs/1709.00962.

73. Z. Zhang, J. M. Girard, Y. Wu, et al., "Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2016), 3438–3446, https://doi.org/10.1109/CVPR.2016.374.

74. J. D. S. Bobbia, R. Macwan, Y. Benezeth, and A. Mansouri, "Unsupervised Skin Tissue Segmentation for Remote Photoplethysmography," *Pattern Recognition Letters* (2017): 82–90.

75. X. Niu, H. Han, S. Shan, and X. Chen, "VIPL-HR: A Multi-Modal Database for Pulse Estimation From Less-Constrained Face Video," *Lecture Notes in Computer Science* 11365 (2019): 562–576, https://doi.org/10.1007/978-3-030-20873-8_36.

76. X. Li, I. Alikhani, J. Shi, et al., "The OBF Database: A Large Face Video Database for Remote Physiological Signal Measurement and Atrial Fibrillation Detection," in *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, (2018), 242–249, https://doi.org/10.1109/FG.2018.00043.

77. Z. Yang, H. Wang, and F. Lu, "Assessment of Deep Learning-Based Heart Rate Estimation Using Remote Photoplethysmography Under Different Illuminations," *IEEE Transactions on Human-Machine Systems* 52, no. 6 (2022): 1236–1246, https://doi.org/10.1109/THMS.2022.3207755.

78. R. M. Sabour, Y. Benezeth, P. De Oliveira, J. Chappé, and F. Yang, "UBFC-phys: A Multimodal Database for Psychophysiological Studies of Social Stress," *IEEE Transactions on Affective Computing* 14, no. 1 (2023): 622–636, https://doi.org/10.1109/TAFFC.2021.3056960.

79. D. McDuff, M. Wander, X. Liu, et al., "SCAMPS: Synthetics for Camera Measurement of Physiological Signals," *Advances in Neural Information Processing Systems* 35 (2022).

80. D. McDuff, "Vital Videos: A Public Dataset of Videos With PPG and BP Ground Truths," *ACM Computing Surveys* 55, no. 9 (2023): 1–40: [Online], https://doi.org/10.1145/3558518.

81. J. Joshi and Y. Cho, "iBVP Dataset: RGB-Thermal rPPG Dataset With High Resolution Signal Quality Labels," *Electronics* 13, no. 7 (2024): 1334, https://doi.org/10.3390/electronics13071334.

82. H. El Boussaki, R. Latif, and A. Saddik, "A Review on Video-Based Heart Rate, Respiratory Rate and Blood Pressure Estimation," in *Advances in Machine Intelligence and Computer Science Applications*, eds. N. Aboutabit, M. Lazaar, and I. Hafidi (Springer Nature Switzerland, 2023), 129–140.

83. A. V. Varsha, C. Markose, and R. P. Aneesh, "Non-Contact Heart Rate Monitoring Using Machine Learning," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies* (ICICICT 2019, 2019), 1400–1404, https://doi.org/10.1109/ICICICT46008.2019.8993251.

84. A. Galli, R. J. H. Montree, S. Que, E. Peri, and R. Vullings, "An Overview of the Sensors for Heart Rate Monitoring Used in Extramural Applications," *Sensors* 22, no. 11 (2022): 1–29, https://doi.org/10.3390/s22114035.

85. S. Tjiharjadi and A. Fajar, "Human Heart Rate Detection Application," in *2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT)*, (2017), 167–172, https://doi.org/10.1109/ICSIIT.2017.12.

86. Y. Akamatsu, T. Umematsu, and H. Imaoka, "CalibrationPhys: Self-Supervised Video-Based Heart and Respiratory Rate Measurements by Calibrating Between Multiple Cameras," *IEEE Journal of Biomedical and Health Informatics* 28, no. 3 (2024): 1460–1471, https://doi.org/10.1109/JBHI.2023.3345486.

87. X. Niu, S. Shan, H. Han, and X. Chen, "RhythmNet: End-to-End Heart Rate Estimation From Face via Spatial-Temporal

Representation," *IEEE Transactions on Image Processing* 29 (2020): 2409–2423, https://doi.org/10.1109/TIP.2019.2947204.

88. F. Yang, S. He, S. Sadanand, A. Yusuf, and M. Bolic, "Contactless Measurement of Vital Signs Using Thermal and RGB Cameras: A Study of COVID-19-Related Health Monitoring," *Sensors* 22, no. 627 (2022): 627, https://doi.org/10.3390/s22020627.

89. F. Shirbani, N. Hui, I. Tan, M. Butlin, and A. P. Avolio, "Effect of Ambient Lighting and Skin Tone on Estimation of Heart Rate and Pulse Transit Time From Video Plethysmography," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (EMBC, 2020), 2642–2645, https://doi.org/10.1109/EMBC44109.2020.9176731.

90. N. Mirabet-Herranz, K. Mallat, and J.-L. Dugelay, "Deep Learning for Remote Heart Rate Estimation: A Reproducible and Optimal State-Of-The-Art Framework BT - Pattern Recognition," in *Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges*, eds. J.-J. Rousseau and B. Kapralos (Springer Nature, 2023), 558–573.

91. A. Caroppo, A. Manni, G. Rescio, P. Siciliano, and A. Leone, "Vital Signs Estimation in Elderly Using Camera-Based Photoplethysmography," *Multimedia Tools and Applications* 83, no. 24 (2024): 65363–65386, https://doi.org/10.1007/s11042-023-18053-3.

92. M. C. Toley, R. G. Mishra, and V. Shirsath, "Facial Video Analytics: An Intelligent Approach to Heart Rate Estimation Using AI Framework," in *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, (2024), 1–7, https://doi.org/10.1109/ESCI59607.2024.10497222.

93. C. Zhao, H. Wang, and Y. Feng, "MSSTNet: Multi-Scale Facial Videos Pulse Extraction Network Based on Separable Spatiotemporal Convolution and Dimension Separable Attention," *Virtual Reality and Intelligent Hardware* 5, no. 2 (2023): 124–141, https://doi.org/10.1016/j.vrih.2022.07.001.

94. C. H. Antink, H. Gao, C. Brüser, and S. Leonhardt, "Beat-to-beat Heart Rate Estimation Fusing Multimodal Video and Sensor Data," *Biomedical Optics Express* 6, no. 8 (2015): 2895, https://doi.org/10.1364/boe.6.002895.

95. A. Anil Jalaja and M. Kavitha, "Contactless Face Video Based Vital Signs Detection Framework for Continuous Health Monitoring Using Feature Optimization and Hybrid Neural Network," *Biotechnology and Bioengineering* 121, no. 4 (2024): 1191–1215, https://doi.org/10.1002/bit.28644.

96. C. H. Cheng, K. L. Wong, J. W. Chin, T. T. Chan, and R. H. Y. So, "Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda," *Sensors* 21, no. 18 (September 2021): 6296, https://doi.org/10.3390/s21186296.

97. R. H. Hamilton and H. K. Davison, "Legal and Ethical Challenges for HR in Machine Learning," *Employee Responsibilities and Rights Journal* 34, no. 1 (2022): 19–39, https://doi.org/10.1007/s10672-021-09377-z.

98. A. Biswal, S. Nanda, C. R. Panigrahi, S. K. Cowlessur, and B. Pati, "Human Activity Recognition Using Machine Learning: A Review," in *Progress in Advanced Computing and Intelligent Engineering*, eds. C. R. Panigrahi, B. Pati, B. K. Pattanayak, S. Amic, and K.-C. Li (Springer Singapore, 2021), 323–333.

99. P. Narayankar and V. P. Baligar, "A Systematic Review of Video Analytics Using Machine Learning and Deep Learning—A Survey," in *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*, eds. M. S. Kaiser, J. Xie, and V. S. Rathore (Springer Nature, 2021), 659–666.

100. D. Qiao, A. H. Ayesha, F. Zulkernine, N. Jaffar, and R. Masroor, "ReViSe: Remote Vital Signs Measurement Using Smartphone Camera," *IEEE Access* 10 (2022): 131656–131670, https://doi.org/10.1109/ACCESS.2022.3229977.

101. V. Selvaraju, N. Spicher, J. Wang, et al., "Continuous Monitoring of Vital Signs Using Cameras: A Systematic Review," *Sensors* 22, no. 11 (June 2022): 4097, https://doi.org/10.3390/s22114097.

102. S. S. Bashar, M. S. Miah, A. H. M. Z. Karim, and M. A. Al Mahmud, "Extraction of Heart Rate From PPG Signal: A Machine Learning Approach Using Decision Tree Regression Algorithm," in *2019 4th International Conference on Electrical Information and Communication Technology, EICT 2019*, Vol. N/A (2019), 1–5, https://doi.org/10.1109/EICT48899.2019.9068845.

103. D. Elhajjar, B. Wallace, A. Law, R. Goubran, and F. Knoefel, "Machine Learning Assessment of Heart Rate Confidence From Video Magnification," in *2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE)*, (2023), 1–6, https://doi.org/10.1109/BIBE60311.2023.00046.

104. M. W. Qiang Zhu, M. Chen, and C.-W. Wong, "Adaptive Multi-Trace Carving Based on Dynamic Programming," in *Adaptive Multi-Trace Carving Based on Dynamic Programming 2018*, (2018), 1174–1189, https://doi.org/10.1109/ACSSC.2018.8645216.

105. X. Liu, J. Fromm, S. Patel, and D. McDuff, "Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement," *arXiv:2006.03790* (2021), https://arxiv.org/abs/2006.03790.

106. C. Zhao, M. Hu, F. Ju, Z. Chen, Y. Li, and Y. Feng, "Convolutional Neural Network With Spatio-Temporal-Channel Attention for Remote Heart Rate Estimation," *Visual Computer* 39, no. 10 (2023): 4767–4785, https://doi.org/10.1007/s00371-022-02624-w.

107. B. Huang, C.-M. Chang, C.-L. Lin, W. Chen, C.-F. Juang, and X. Wu, "Visual Heart Rate Estimation From Facial Video Based on CNN," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, (2020), 1658–1662, https://doi.org/10.1109/ICIEA48937.2020.9248356.

108. C. Zhao, M. Zhou, W. Han, and Y. Feng, "Anti-Motion Remote Measurement of Heart Rate Based on Region Proposal Generation and Multi-Scale ROI Fusion," *IEEE Transactions on Instrumentation and Measurement* 71 (2022): 1–13, https://doi.org/10.1109/TIM.2022.3169567.

109. Y. Akamatsu, Y. Onishi, and H. Imaoka, "Heart Rate and Oxygen Saturation Estimation From Facial Video With Multimodal Physiological Data Generation," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, (2022), 1111–1115, https://doi.org/10.1109/ICASSP43922.2022.9747109.

110. M. Hu, F. Qian, X. Wang, L. He, D. Guo, and F. Ren, "Robust Heart Rate Estimation With Spatial–Temporal Attention Network From Facial Videos," *IEEE Transactions on Cognitive and Developmental Systems* 14, no. 2 (2022): 639–647, https://doi.org/10.1109/TCDS.2021.3062370.

111. G. Yu, Z.. Shen, Y. Shi, et al., "PhysFormer++: Facial Video-Based Physiological Measurement With SlowFast Temporal Difference Transformer," *International Journal of Computer Vision* 131, no. 6 (2023): 1307–1330, https://doi.org/10.1007/s11263-023-01758-1.

112. K. Kurihara, Y. Maeda, D. Sugimura, and T. Hamamoto, "Spatio-Temporal Structure Extraction of Blood Volume Pulse Using Dynamic Mode Decomposition for Heart Rate Estimation," *IEEE Access* 11, (2023): 59081–59096, https://doi.org/10.1109/ACCESS.2023.3284465.

113. B. Li, P. Zhang, J. Peng, and H. Fu, "Non-Contact PPG signal and Heart Rate Estimation With Multi-Hierarchical Convolutional Network," *Pattern Recognition* 139 (2023): 109421, https://doi.org/10.1016/j.patcog.2023.109421.

114. C. Zhao, H. Wang, H. Chen, W. Shi, and Y. Feng, "JAMSNet: A Remote Pulse Extraction Network Based on Joint Attention and Multi-Scale Fusion," *IEEE Transactions on Circuits and Systems for Video Technology* 33, no. 6 (2023): 2783–2797, https://doi.org/10.1109/TCSVT.2022.3227348.

115. D. M. Huang, J. Huang, K. Qiao, N. S. Zhong, H. Z. Lu, and W. J. Wang, "Deep Learning-Based Lung Sound Analysis for Intelligent

Stethoscope," *Military Medical Research* 10, no. 1 (2023): 1–23, https://doi.org/10.1186/s40779-023-00479-3.

116. A. Di Credico, D. Perpetuini, P. Izzicupo, et al., "Estimation of Heart Rate Variability Parameters by Machine Learning Approaches Applied to Facial Infrared Thermal Imaging," *Frontiers in Cardiovascular Medicine* 9 (May 2022): 1–11, https://doi.org/10.3389/fcvm.2022.893374.

117. F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in Public Health* 5 (September 2017): 1–17, https://doi.org/10.3389/fpubh.2017.00258.

118. E. Mejía-Mejía, K. Budidha, T. Y. Abay, J. M. May, and P. A. Kyriacou, "Heart Rate Variability (HRV) and Pulse Rate Variability (PRV) for the Assessment of Autonomic Responses," *Frontiers in Physiology* 11 (July 2020), https://doi.org/10.3389/fphys.2020.00779.

119. H. Kuang, F. Lv, X. Ma, and X. Liu, "Efficient Spatiotemporal Attention Network for Remote Heart Rate Variability Analysis," *Sensors* 22, no. 3 (2022): 1010, https://doi.org/10.3390/s22031010.

120. U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart Rate Variability: A Review," *Medical, & Biological Engineering & Computing* 44, no. 12 (2006): 1031–1051, https://doi.org/10.1007/s11517-006-0119-0.

121. R. McCraty, B. Barrios-Choplin, D. Rozman, M. Atkinson, and A. D. Watkins, "The Impact of a New Emotional Self-Management Program on Stress, Emotions, Heart Rate Variability, DHEA and Cortisol," *Integrative Physiological and Behavioral Science* 33, no. 2 (1998): 151–170, https://doi.org/10.1007/BF02688660.

122. T. Shaik, X. Tao, N. Higgins, et al., "Remote Patient Monitoring Using Artificial Intelligence: Current State, Applications, and Challenges," *WIREs Data Mining and Knowledge Discovery* 13, no. 2 (March 2023), https://doi.org/10.1002/widm.1485.

123. A. Gudi, M. Bittner, R. Lochmans, and J. Van Gemert, "Efficient Real-Time Camera Based Estimation of Heart Rate and Its Variability," in *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, (2019), 1570–1579, https://doi.org/10.1109/ICCVW.2019.00196.

124. J.-G. Dong, "The Role of Heart Rate Variability In Sports Physiology," *Experimental and Therapeutic Medicine* 11, no. 5 (May 2016): 1531–1536, https://doi.org/10.3892/etm.2016.3104.

125. Z. Yu, X. Li, and G. Zhao, "Recovering Remote Photoplethysmograph Signal From Facial Videos Using Spatio-Temporal Convolutional Networks," *ArXiv* abs/1905.0 (2019): [Online], https://api.semanticscholar.org/CorpusID:146808348.

126. H. Lu, H. Han, and S. K. Zhou, "DuAl-GaN: Joint BVP and Noise Modeling for Remote Physiological Measurement," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2021), 12399–12408, https://doi.org/10.1109/CVPR46437.2021.01222.

127. S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, no. 1 (2013): 221–231, https://doi.org/10.1109/TPAMI.2012.59.

128. T. Luguev, D. Seuß, and J.-U. Garbas, "Deep Learning based Affective Sensing With Remote Photoplethysmography," in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, (2020), 1–4, https://doi.org/10.1109/CISS48834.2020.1570617362.

129. R. Song, H. Chen, J. Cheng, C. Li, Y. Liu, and X. Chen, "PulseGAN: Learning to Generate Realistic Pulse Waveforms in Remote Photoplethysmography," *IEEE Journal of Biomedical and Health Informatics* 25, no. 5 (2021): 1373–1384, https://doi.org/10.1109/JBHI.2021.3051176.

130. G. H. Martinez-Delgado, A. J. Correa-Balan, J. A. May-Chan, C. E. Parra-Elizondo, L. A. Guzman-Rangel, and A. Martinez-Torteya, "Measuring Heart Rate Variability Using Facial Video," *Sensors* 22, no. 13 (June 2022): 4690, https://doi.org/10.3390/s22134690.

131. C. A. Elliott M, M. Elliott, and A. Coventry, "Critical Care: The Eight Vital Signs of Patient Monitoring," *British Journal of Nursing* 21, no. 10 (May 2012): 621–625, https://doi.org/10.12968/bjon.2012.21.10.621.

132. D. Titisari and M. P. A. T. Putra, "Assessing the Effectiveness of Mechanical Sensors for Respiratory Rate Detection BT," in *Proceedings of the 4th International Conference on Electronics, Biomedical Engineering, and Health Informatics*, eds. T. Triwiyanto, A. Rizal, and W. Caesarendra (Springer Nature, 2024), 519–530.

133. A. Nicolò, C. Massaroni, E. Schena, and M. Sacchetti, "The Importance of Respiratory Rate Monitoring: From Healthcare to Sport And Exercise," *Sensors* 20, no. 21 (2020): 1–45, https://doi.org/10.3390/s20216396.

134. R. B. Schlesinger, *Respiratory Rate BT - Dictionary of Toxicology*, ed. A. B. Pant (Springer Nature, 2024), 882, https://doi.org/10.1007/978-981-99-9283-6_2382.

135. B. T. Thompson, R. C. Chambers, and K. D. Liu, "Acute Respiratory Distress Syndrome," *New England Journal of Medicine* 377, no. 6 (August 2017): 562–572, https://doi.org/10.1056/NEJMra1608077.

136. Y. C. Lee, A. Syakura, M. A. Khalil, C. H. Wu, Y. F. Ding, and C. W. Wang, "A Real-Time Camera-Based Adaptive Breathing Monitoring System," *Medical, & Biological Engineering & Computing* 59, no. 6 (2021): 1285–1298, https://doi.org/10.1007/s11517-021-02371-5.

137. J. H. Lee, L. A. Nathanson, R. C. Burke, B. W. Anthony, N. I. Shapiro, and A. S. Dagan, "Assessment of Respiratory Rate Monitoring in the Emergency Department," *JACEP Open* 5, no. 3 (June 2024): e13154, https://doi.org/10.1002/emp2.13154.

138. P. E. Marik, "Understanding the Vital Signs: BP, HR, RR, TEMP, SaO2 and SV," in *Evidence-Based Critical Care* (Springer International Publishing, 2015), 169–196, https://doi.org/10.1007/978-3-319-11020-2_14.

139. C. Massaroni, G. Senesi, E. Schena, and S. Silvestri, "Analysis of Breathing via Optoelectronic Systems: Comparison of Four Methods for Computing Breathing Volumes and Thoraco-Abdominal Motion Pattern," *Computer Methods in Biomechanics and Biomedical Engineering* 20, no. 16 (2017): 1678–1689, https://doi.org/10.1080/10255842.2017.1406081.

140. A. Qayyum, M. K. A. Ahamed Khan, M. Mazher, M. Suresh, D. N. Jamal, and J. T. Duc Chung, "Convolutional Neural Network Approach for Estimating Physiological States Involving Face Analytics," in *2019 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2019 - Proceedings*, (2019), 68–72, https://doi.org/10.1109/I2CACIS.2019.8825048.

141. N. S. Suriani, N. S. Shahdan, N. M. Sahar, and N. S. A. M. Taujuddin, "Non-Contact Facial Based Vital Sign Estimation Using Convolutional Neural Network Approach," *International Journal of Advanced Computer Science and Applications* 13, no. 5 (2022): 386–393, https://doi.org/10.14569/IJACSA.2022.0130546.

142. M. Gwak, K. Vatanparvar, L. Zhu, and N. Rashid, "Multimedia Prattng Rat Enjoy Asing In Every Mission With Your Lord From Rajab Samra," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2024), 2046–2050, https://doi.org/10.1109/ICASSP48485.2024.10446086.

143. L. Antognoli, P. Marchionni, S. Spinsante, S. Nobile, V. P. Carnielli, and L. Scalise, "Enanced Video Heart Rate and Respiratory Rate Evaluation: Standard Multiparameter Monitor vs Clinical Confrontation in Newborn Patients," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, (2019), 1–5, https://doi.org/10.1109/MeMeA.2019.8802147.

144. H. K. M. S. G. Negishi, T., A. Shigeto, Matsui, T., Liu, et al., "Contactless Vital Signs Measurement System Using Rgb-Thermal

Image Sensors And Its Clinical Screening Test On Patients With Seasonal Influenza," *Sensors* 20, no. 8 (2020): 2171, https://doi.org/10.3390/s20082171.

145. Z. M. Lazri, Q. Zhu, M. Chen, M. Wu, and Q. Wang, "Detecting Essential Landmarks Directly in Thermal Images for Remote Body Temperature and Respiratory Rate Measurement With a Two-Phase System," *IEEE Access* 10 (2022): 39080–39094, https://doi.org/10.1109/ACCESS.2022.3161968.

146. G. Sun, Y. Nakayama, S. Dagdanpurev, et al., "Remote Sensing of Multiple Vital Signs Using A Cmos Camera-Equipped Infrared Thermography System and Its Clinical Application In Rapidly Screening Patients With Suspected Infectious Diseases," *International Journal of Infectious Diseases* 55 (2017): 113–117, https://doi.org/10.1016/j.ijid.2017.01.007.

147. T. Negishi, G. Sun, H. Liu, S. Sato, T. Matsui, and T. Kirimoto, "Stable Contactless Sensing of Vital Signs Using RGB-Thermal Image Fusion System With Facial Tracking for Infection Screening," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBS, 2018), 4371–4374, https://doi.org/10.1109/EMBC.2018.8513300.

148. Y. Cho, S. J. Julier, N. Marquardt, and N. Bianchi-Berthouze, "Robust Tracking of Respiratory Rate in High-Dynamic Range Scenes Using Mobile Thermal Imaging," *Biomedical Optics Express* 8, no. 10 (October 2017): 4480–4503, https://doi.org/10.1364/BOE.8.004480.

149. M. Gwak, K. Vatanparvar, L. Zhu, J. Kuang, and A. Gao, "Contactless Monitoring of Respiratory Rate and Breathing Absence From Head Movements Using an RGB Camera," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE, 2023), 1–4.

150. Y. Ren, B. Syrnyk, and N. Avadhanam, "Improving Video-Based Heart Rate and Respiratory Rate Estimation via Pulse-Respiration Quotient," *Proceedings of Machine Learning Research* 184, no. (July 2022): 136–145.

151. A. De Groote, M. Wantier, G. Cheron, M. Estenne, and M. Paiva, "Chest Wall Motion During Tidal Breathing," *Journal of Applied Physiology* 83, no. 5 (1997): 1531–1537, https://doi.org/10.1152/jappl.1997.83.5.1531.

152. K. Poorzargar, C. Pham, D. Panesar, et al., "Video Plethysmography for Contactless Measurement of Respiratory Rate in Surgical Patients," *Journal of Clinical Monitoring and Computing* 38, no. 1 (2024): 47–55, https://doi.org/10.1007/s10877-023-01064-8.

153. F. Schrumpf, C. Monch, G. Bausch, and M. Fuchs, "Exploiting Weak Head Movements for Camera-Based Respiration Detection," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBS, 2019), 6059–6062, https://doi.org/10.1109/EMBC.2019.8856387.

154. M. Gwak, K. Vatanparvar, J. Kuang, and A. Gao, "Motion-Based Respiratory Rate Estimation With Motion Artifact Removal Using Video of Face and Upper Body," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE, 2022), 1961–1967.

155. M. Pediaditis, C. Farmaki, S. Schiza, N. Tzanakis, E. Galanakis, and V. Sakkalis, "Contactless Respiratory Rate Estimation From Video in a Real-Life Clinical Environment Using Eulerian Magnification and 3d Cnns," in *IST 2022 - IEEE International Conference on Imaging Systems and Techniques, Proceedings*, (2022), 1–6, https://doi.org/10.1109/IST55454.2022.9827675.

156. M. Ghodratigohar, H. Ghanadian, and H. Al Osman, "A Remote Respiration Rate Measurement Method for Non-Stationary Subjects Using CEEMDAN and Machine Learning," *IEEE Sensors Journal* 20, no. 3 (2020): 1400–1410, https://doi.org/10.1109/JSEN.2019.2946132.

157. S. Lyra, L. Mayer, L. Ou, et al., "A Deep Learning-Based Camera Approach for Vital Sign Monitoring Using Thermography Images for ICU Patients," *Sensors* 21, no. 4 (2021): 1–18, https://doi.org/10.3390/s21041495.

158. Q. Zhan, J. Hu, Z. Yu, X. Li, and W. Wang, "Revisiting Motion-Based Respiration Measurement From Videos," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBS, 2020), 5909–5912, https://doi.org/10.1109/EMBC44109.2020.9175662.

159. A. Kwasniewska, M. Szankin, J. Ruminski, and M. Kaczmarek, "Evaluating Accuracy of Respiratory Rate Estimation From Super Resolved Thermal Imagery," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBS, 2019), 2744–2747, https://doi.org/10.1109/EMBC.2019.8857764.

160. P. Jagadev, S. Naik, and L. I. Giri, "Contactless Monitoring of Human Respiration Using Infrared Thermography and Deep Learning," *Physiological Measurement* 43, no. 2 (March 2022): 25006, https://doi.org/10.1088/1361-6579/ac57a8.

161. J. Brieva, H. Ponce, and E. Moya-Albor, "Non-Contact Breathing Rate Estimation Using Machine Learning With an Optimized Architecture," *Mathematics* 11, no. 3 (2023): 645, https://doi.org/10.3390/math11030645.

162. W. Chen and D. McDuff, "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks," *Lecture Notes in Computer Science* 11206 (2018): 356–373, https://doi.org/10.1007/978-3-030-01216-8_22.

163. P. Villarroel, M.، Chaichulee, S.، Jorge, et al., "Non-Contact Physiological Monitoring of Preterm Infants in the Neonatal Intensive Care Unit," *Npj Digital Medicine* 2, no. 1 (2019): 1–18, https://doi.org/10.1038/s41746-019-0199-5.

164. T. Negishi, S. Abe, T. Matsui, et al., "Contactless Vital Signs Measurement System Using RGB-Thermal Image Sensors and Its Clinical Screening Test on Patients With Seasonal Influenza," *Sensors* 20, no. 8 (2020): 2171, https://doi.org/10.3390/s20082171.

165. J. Laurie, N. Higgins, T. Peynot, L. Fawcett, and J. Roberts, "An Evaluation of a Video Magnification-Based System For Respiratory Rate Monitoring in an Acute Mental Health Setting," *International Journal of Medical Informatics* 148 (April 2021): 104378, https://doi.org/10.1016/j.ijmedinf.2021.104378.

166. Z. Sun and X. Li, "Contrast-Phys: Unsupervised Video-Based Remote Physiological Measurement via Spatiotemporal Contrast," *Lecture Notes in Computer Science* 13672 (2022): 492–510, https://doi.org/10.1007/978-3-031-19775-8_29.

167. M. Alnaggar, A. I. Siam, M. Handosa, T. Medhat, and M. Z. Rashad, "Video-Based Real-Time Monitoring for Heart Rate and Respiration Rate," *Expert Systems With Applications* 225 (2023): 120135, https://doi.org/10.1016/j.eswa.2023.120135.

168. R. H. Goudarzi, S. Somayyeh Mousavi, and M. Charmi, "Using Imaging Photoplethysmography (iPPG) Signal for Blood Pressure Estimation," in *2020 International Conference on Machine Vision and Image Processing (MVIP)*, (2020), 1–6, https://doi.org/10.1109/MVIP49855.2020.9116902.

169. R. N. Haldar, "Global Brief on Hypertension: Silent Killer, Global Public Health Crisis," *Indian Journal of Physical Medicine and Rehabilitation* 24, no. 1 (2013): 2, https://doi.org/10.5005/ijopmr-24-1-2.

170. S. Baek, J. Jang, and S. Yoon, "End-to-End Blood Pressure Prediction via Fully Convolutional Networks," *IEEE Access* 7 (2019): 185458–185468, https://doi.org/10.1109/ACCESS.2019.2960844.

171. Y. Zhou, H. Ni, Q. Zhang, and Q. Wu, "The Noninvasive Blood Pressure Measurement Based on Facial Images Processing," *IEEE Sensors Journal* 19, no. 22 (2019): 10624–10634, https://doi.org/10.1109/JSEN.2019.2931775.

172. C. Gonzalez Viejo, S. Fuentes, D. D. Torrico, and F. R. Dunshea, "Non-Contact Heart Rate and Blood Pressure Estimations From Video

Analysis and Machine Learning Modelling Applied to Food Sensory Responses: A Case Study for Chocolate," *Sensors* 18, no. 6 (June 2018): 1802, https://doi.org/10.3390/s18061802.

173. H. Luo, D. Yang, A. Barszczyk, et al., "Smartphone-Based Blood Pressure Measurement Using Transdermal Optical Imaging Technology," *Circulation: Cardiovascular Imaging* 12, no. 8 (2019): 1–10, https://doi.org/10.1161/CIRCIMAGING.119.008857.

174. Y. Gao, O. R. Patil, B. Li, and Z. Jin, "CamBP: A Camera-Based, Non-Contact Blood Pressure Monitor," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (October 2018), https://doi.org/10.1145/3123024.3124428.

175. Q. V. Tran, S. F. Su, Q. M. Tran, and V. Truong, "Intelligent Non-Invasive Vital Signs Estimation From Image Analysis," in *2020 International Conference on System Science and Engineering, ICSSE 2020*, (2020), https://doi.org/10.1109/ICSSE50014.2020.9219297.

176. P. Yawle and R. B. Ghongade, "ML-Based Blood Pressure Estimation Using Converted PPG Signal From Video," in *Proceedings of the 18th INDIAcom; 2024 11th International Conference on Computing for Sustainable Global Development, INDIACom*, (2024), 1182–1187, https://doi.org/10.23919/INDIACom61295.2024.10498176.

177. W. Chen, D. Zhai, H. Wu, et al., "Non-Contact Blood Pressure Detection Based on Weighted Ensemble Learning Model," *Signal Image Video Process* 18, no. 1 (2024): 553–560, https://doi.org/10.1007/s11760-023-02762-1.

178. D. Djeldjli, F. Bousefsaf, C. Maaoui, F. Bereksi-Reguig, and A. Pruski, "Remote Estimation of Pulse Wave Features Related to Arterial Stiffness and Blood Pressure Using a Camera," *Biomedical Signal Processing and Control* 64 (January 2021): 102242, https://doi.org/10.1016/j.bspc.2020.102242.

179. J. Speth, N. Vance, B. Sporrer, L. Niu, P. Flynn, and A. Czajka, "MSPM: A Multi-Site Physiological Monitoring Dataset for Remote Pulse, Respiration, and Blood Pressure Estimation," *IEEE Transactions on Instrumentation and Measurement* 73 (2024): 1–15: [Online], https://arxiv.org/abs/2402.02224v1.

180. B. F. Wu, L. W. Chiu, Y. C. Wu, C. C. Lai, H. M. Cheng, and P. H. Chu, "Contactless Blood Pressure Measurement Via Remote Photoplethysmography With Synthetic Data Generation Using Generative Adversarial Networks," *IEEE Journal of Biomedical and Health Informatics* 28, no. 2 (2024): 621–632, https://doi.org/10.1109/JBHI.2023.3265857.

181. B.-F. Wu, L.-W. Chiu, Y.-C. Wu, C.-C. Lai, and P.-H. Chu, "Contactless Blood Pressure Measurement via Remote Photoplethysmography With Synthetic Data Generation Using Generative Adversarial Network," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2022), 2129–2137, https://doi.org/10.1109/CVPRW56347.2022.00231.

182. K. Iuchi, R. Miyazaki, G. C. Cardoso, K. Ogawa-Ochiai, and N. Tsumura, "Remote Estimation of Continuous Blood Pressure by a Convolutional Neural Network Trained on Spatial Patterns of Facial Pulse Waves," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2022), 2138–2144, https://doi.org/10.1109/CVPRW56347.2022.00232.

183. C. Qin, Y. Li, C. Liu, and X. Ma, "Cuff-Less Blood Pressure Prediction Based on Photoplethysmography and Modified ResNet," *Bioengineering* 10, no. 4 (2023): 1–14, https://doi.org/10.3390/bioengineering10040400.

184. H. A. S. Nichole, R. Daniela, A. Y. Molla, et al., "Pulse Oximetry in Low-Resource Settings During the COVID-19 Pandemic," *Lancet Global Health* 8, no. 9 (September 2020): e1121–e1122, https://doi.org/10.1016/S2214-109X(20)30287-4.

185. M. K. Lingqin, Y. Zhao, L. Dong, et al., "Non-Contact Detection of Oxygen Saturation Based on Visible Light Imaging Device Using Ambient Light," *Optics Express* 21, no. 15 (2013): 17464, https://doi.org/10.1364/oe.21.017464.

186. A. Al-Naji, G. A. Khalid, J. F. Mahdi, and J. Chahl, "Non-Contact spo2 Prediction System Based on a Digital Camera," *Applied Sciences* 11, no. 9 (2021): 4255, https://doi.org/10.3390/app11094255.

187. X. Ding, D. Nassehi, and E. C. Larson, "Measuring Oxygen Saturation With Smartphone Cameras Using Convolutional Neural Networks," *IEEEJournal of Biomedical and Health Informatics* 23, no. 6 (2019): 2603–2610, https://doi.org/10.1109/JBHI.2018.2887209.

188. J. Y. A. Foo, K. P. Chua, and X. J. A. Tan, "Clinical Applications and Issues of Oxygen Saturation Level Measurements Obtained From Peripheral Sites," *IEEE Journal of Biomedical and Health Informatics* 37, no. 6 (2013): 388–395, https://doi.org/10.3109/03091902.2013.816380.

189. J. Mathew, X. Tian, C. W. Wong, S. Ho, D. K. Milton, and M. Wu, "Remote Blood Oxygen Estimation From Videos Using Neural Networks," *IEEE Journal of Biomedical and Health Informatics* 27, no. 8 (2023): 3710–3720, https://doi.org/10.1109/JBHI.2023.3236631.

190. D. Dias and J. P. S. Cunha, "Wearable Health Devices—Vital Sign Monitoring, Systems and Technologies," *Sensors* 18, no. 8 (2018): 2414, https://doi.org/10.3390/s18082414.

191. L. Guazzi, A. R. Villarroel, M. Jorge, et al., "Non-Contact Measurement of Oxygen Saturation With an RGB Camera," *Biomedical Optics Express* 6, no. 9 (2015): 3320, https://doi.org/10.1364/boe.6.003320.

192. L. M. O'Brien, V. A. Stebbens, C. F. Poets, E. G. Heycock, and D. P. Southall, "Oxygen Saturation During the First 24 Hours of life," *Archives of Disease in Childhood - Fetal and Neonatal Edition* 83, no. 1 (July 2000): F35 LP-F38, https://doi.org/10.1136/fn.83.1.F35.

193. T. Stogiannopoulos, G.-A. Cheimariotis, and N. Mitianoudis, "A Non-Contact SpO2 Estimation Using Video Magnification and Infrared Data," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2023), 1–5, https://doi.org/10.1109/ICASSP49357.2023.10095244.

194. D. Shao, C. Liu, F. Tsow, et al., "Noncontact Monitoring of Blood Oxygen Saturation Using Camera and Dual-Wavelength Imaging System," *IEEE Transactions on Biomedical Engineering* 63, no. 6 (2016): 1091–1098, https://doi.org/10.1109/TBME.2015.2481896.

195. B. B. Hafen and S. Sharma, "Oxygen Saturation," in *StatPearls*, (StatPearls Publishing, 2025), https://www.ncbi.nlm.nih.gov/books/NBK541123/.

196. G. Casalino, G. Castellano, and G. Zaza, "A mHealth Solution for Contact-Less Self-Monitoring of Blood Oxygen Saturation," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, (2020), 1–7, https://doi.org/10.1109/ISCC50000.2020.9219718.

197. Y. Akamatsu, Y. Onishi, and H. Imaoka, "Heart Rate and Oxygen Saturation Estimation From Facial Video With Multimodal Physiological Data Generation," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2022), 1111–1115, https://doi.org/10.1109/ICASSP43922.2022.9747109.

198. J. C. Cobos-Torres and M. Abderrahim, "Simple Measurement of Pulse Oximetry Using a Standard Color Camera," in *2017 40th International Conference on Telecommunications and Signal Processing, TSP 2017*, (2017), 452–455, https://doi.org/10.1109/TSP.2017.8076026.

199. J. Peng, W. Su, H. Chen, J. Sun, and Z. Tian, "CL-SPO2Net: Contrastive Learning Spatiotemporal Attention Network for Non-Contact Video-Based SpO2 Estimation," *Bioengineering* 11, no. 2 (2024): 113, https://doi.org/10.3390/bioengineering11020113.

200. B.-J. Wu, B.-F. Wu, Y.-C. Dong, H.-C. Lin, and P.-H. Li, "Peripheral Oxygen Saturation Measurement Using an RGB camera," *IEEE Sensors Journal* 23, no. 21 (2023): 26551–26563, https://doi.org/10.1109/jsen.2023.3284196.

201. M. Yoshizawa, N. Sugita, A. Tanaka, A. Togashi, I. Kaji, and T. Yambe, "Basic Approach to Estimation of Blood Oxygen Saturation

Using an RGB Color Camera Without Infrared Light," in *2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech)*, (2022), 68–71, https://doi.org/10.1109/LifeTech53646.2022.9754752.

202. "Absorption Spectra of Hemoglobin," [Online], Oregon Medical Laser Center, accessed June 17, 2025, https://omlc.org/spectra/hemoglobin/.

203. X. Tian, C.-W. Wong, S. M. Ranadive, and M. Wu, "A Multi-Channel Ratio-of-Ratios Method for Noncontact Hand Video Based SpO$_2$ Monitoring Using Smartphone Cameras," *IEEE Journal of Selected Topics in Signal Processing* 16, no. 2 (2022): 197–207, https://doi.org/10.1109/JSTSP.2022.3152352.

204. L. Zhu, K. Vatanparvar, M. Gwak, J. Kuang, and A. Gao, "Contactless SpO2 Detection From Face Using Consumer Camera," in *2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, (2022), 1–4, https://doi.org/10.1109/BSN56160.2022.9928509.

205. B. Hamoud, W. Othman, N. Shilov, and A. Kashevnik, "Contactless Oxygen Saturation Detection Based on Face Analysis: An Approach and Case Study," in *2023 33rd Conference of Open Innovations Association (FRUCT)*, (2023), 54–62, https://doi.org/10.23919/FRUCT58615.2023.10143059.

206. M. Hu, X. Wu, X. Wang, Y. Xing, N. An, and P. Shi, "Contactless Blood Oxygen Estimation From Face Videos: A Multi-Model Fusion Method Based on Deep Learning," *Biomedical Signal Processing and Control* 81 (2023): 104487, https://doi.org/10.1016/j.bspc.2022.104487.

207. C. H. Cheng, Z. Yuen, S. Chen, et al., "Contactless Blood Oxygen Saturation Estimation From Facial Videos Using Deep Learning," *Bioengineering* 11, no. 3 (2024): 1–16, https://doi.org/10.3390/bioengineering11030251.

208. Y. Akamatsu, Y. Onishi, and H. Imaoka, "Blood Oxygen Saturation Estimation From Facial Video Via DC and AC Components of Spatio-Temporal Map," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2023), 1–5, https://doi.org/10.1109/ICASSP49357.2023.10096616.

209. X. Sun, T. Wen, W. Chen, and B. Huang, "CCSpO2Net: Camera-Based Contactless Oxygen Saturation Measurement Foundation Model in Clinical Settings," *IEEE Transactions on Instrumentation and Measurement* 73 (2024): 1–11, https://doi.org/10.1109/TIM.2024.3375409.

210. B. Huang, S. Hu, Z. Liu, et al., "Challenges and Prospects of Visual Contactless Physiological Monitoring in Clinical Study," *Nature Research* 6, no. 1 (December 2023): 231, https://doi.org/10.1038/s41746-023-00973-x.

211. Y. Chu, K. Tang, Y. C. Hsu, et al., "Non-Invasive Arterial Blood Pressure Measurement and SpO2 Estimation Using PPG Signal: A Deep Learning Framework," *BMC Medical Informatics and Decision Making* 23, no. 1 (2023): 131, https://doi.org/10.1186/s12911-023-02215-2.

212. A. R. James S, E. Ward, H. Lacroix, et al., "Hypothermia Therapy After Traumatic Brain Injury in Children," *New England Journal of Medicine* 358, no. 23 (2008): 2447–2456, https://doi.org/10.1056/NEJMoa0706930.

213. N. Sellier, E. Guettier, and C. Staub, "A Review of Methods to Measure Animal Body Temperature in Precision Farming," (2014) [Online], https://api.semanticscholar.org/CorpusID:134181373.

214. Z. J. Yao K, Y. Chen, Z. D. Zhang, G. Li, J. B. Zhai, and Y. H. Liu, "Identity and Body Temperature Detection System Based on Image Registration," *Journal of Physics: Conference Series* 2290, no. 1 (June 2022): 012073, https://doi.org/10.1088/1742-6596/2290/1/012073.

215. C. Lyra, S.، Mayer, L.، Ou, et al., "A Deep Learning-Based Camera Approach for Vital Sign Monitoring Using Thermography Images for ICU Patients," *Sensors* 21, no. 4 (2021): 1495, https://doi.org/10.3390/s21041495.

216. S. Dagdanpurev, G. Sun, L. Choimaa, S. Abe, and T. Matsui, "Clinical Application of Multiple Vital Signs-Based Infection Screening System in a Mongolian Hospital: Optimization of Facial Temperature Measurement by Thermography at Various Ambient Temperature Conditions Using Linear Regression Analysis," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2018), 5313–5316.

217. L. Unursaikhan, B. Amarsanaa, G. Sun, et al., "Development of a Novel Vital-Signs-Based Infection Screening Composite-Type Camera With Truncus Motion Removal Algorithm to Detect COVID-19 Within 10 Seconds and Its Clinical Validation," *Frontiers in Physiology* 13 (2022): 1–10, https://doi.org/10.3389/fphys.2022.905931.

218. S. Lyra, J. Rixen, K. Heimann, et al., "Camera Fusion for Real-Time Temperature Monitoring of Neonates Using Deep Learning," *Medical, & Biological Engineering & Computing* 60, no. 6 (2022): 1787–1800, https://doi.org/10.1007/s11517-022-02561-9.

219. A. Dasari, A. Revanur, L. A. Jeni, and C. S. Tucker, "Video-Based Elevated Skin Temperature Detection," *IEEE Transactions on Biomedical Engineering* 70, no. 8 (2023): 2430–2444, https://doi.org/10.1109/TBME.2023.3247910.

220. P. Jakkaew and T. Onoye, "Non-Contact Respiration Monitoring and Body Movements Detection for Sleep Using Thermal Imaging," *Sensors* 20, no. 21 (2020): 1–14, https://doi.org/10.3390/s20216307.

221. Y. Zheng, H. Wang, and Y. Hao, "Mobile Application for Monitoring Body Temperature From Facial Images Using Convolutional Neural Network and Support Vector Machine," in *Mobile Multimedia/Image Processing, Security, and Applications 2020*, eds. S. S. Agaian, V. K. Asari, S. P. DelMarco, and S. A. Jassim (SPIE, 2020), 113990B, https://doi.org/10.1117/12.2557856.