Chapter 3
Facial Tracking and Motion Amplification for Capturing Heartbeats in Video

Yuri Alencar, Camilly Costa, Gustavo Castro, Silvana Oliveira, Victor Cardoso, Moisés Silva, and João Weyl

Abstract Motion magnification is a technique that amplifies imperceptible variations in videos, highlighting subtle movements and even color changes associated with physiological signals. This approach has proven effective for non-invasive monitoring of physiological signals, such as heart rate and respiratory rate, by amplifying subtle motion oscillations linked to blood flow. The technique typically integrates Principal Component Analysis (PCA) and Blind Source Separation (BSS) to remove redundant information and isolate meaningful physiological signals. On the other hand, the Hilbert motion magnification technique offers greater precision in capturing high-frequency, low-amplitude oscillations. However, directly extracting signals from raw video can be compromised by background noise, lighting variations, and unwanted global motion. The implementation of MediaPipe's technique addresses these challenges by accurately isolating and tracking the anatomical regions of interest on a person's face. Continuous tracking of these facial points stabilizes head position, removing unwanted rigid movements before motion decomposition. Dynamically selecting this sub-image significantly reduces the volume of processed data and improves the signal-to-noise ratio, as motion extraction is applied only to relevant pixels, avoiding artifacts from camera movement or vibrations. This preprocessing further enhances the performance of the motion magnification framework, enabling it to more accurately isolate the cardiac component from other independent sources in the amplified signal. The proposed method shows promise for applications in telemedicine, neonatal monitoring, and clinical studies, paving the way for remote, scalable, and contactless healthcare systems due to its automation and low computational cost.

Keywords Motion magnification · Facial tracking c· Heart-rate monitorin · Non-contact monitoring · Blind source separation

Introduction

Heart rate (HR) monitoring is a fundamental procedure for analyzing cardiovascular condition and tracking patients in different health contexts. Its relevance is notable in fields ranging from physical training to stress management. HR acts as a thermometer for the body's reactions to various stimuli, such as physical exertion, psychological stress, or illnesses [1]. Current research reinforces that specific metrics, such as resting heart rate (RHR), act as a screening tool for cardiometabolic risk, with chronically high values potentially indicating an autonomic imbalance [1]. Similarly, the speed at which the heart slows down after an activity is a strong indicator of physical fitness and heart health [2]. In this context, heart rate variability (HRV) stands out as a highly relevant marker, as it mirrors the autonomic nervous system's ability to modulate the heart rhythm in different scenarios. High HRV is often interpreted as a sign of good cardiovascular health and an efficient capacity to adapt to stress. Conversely, decreased HRV may suggest autonomic dysfunction, serving as an objective biomarker for stress states and being associated with various pathologies, especially those of cardiovascular origin [3]. Despite its vast utility, HR

Yuri Alencar · Camilly Costa · Gustavo Castro · Silvana Oliveira · Victor Cardoso · João Weyl
Laboratório de Eletromagnetismo Aplicado, Universidade Federal do Pará, Belém, PA 66075, Brazil
e-mail: yuri.alencar@itec.ufpa.br; caam.cost@gmail.com; gustavonascimentocastro1@gmail.com; silvana.oliveira@itec.ufpa.br; victorcard@ufpa.br; mfelipe@lanl.gov

Moisés Silva
Los Alamos National Laboratory, Los Alamos, NM 87545
e-mail: joao.weyl@gmail.com

Kristin B. Zimmerman, PhD (eds.), *Structural Health Monitoring, Metamaterials, Metastructures and Additive Manufacturing and Dynamics of Biomedical and Bioinspired Systems, 2026, Vol. 10*, Conference Proceedings
of the Society for Experimental Mechanics Series, https://doi.org/10.1007/978-8-743-81449-8_3

measurement methods, especially the most modern and accessible ones, are not without challenges. Photoplethysmography (PPG) technology, the basis of most wearable devices, although very practical, can have its accuracy affected by several factors. The main source of errors is motion artifacts, which occur during exercise and can generate incorrect readings [4]. Additionally, individual characteristics such as the presence of tattoos at the sensor site or low blood perfusion (common in cold environments) can weaken the signal and compromise data accuracy. These technical limitations show that for medical diagnoses or for athletes seeking maximum precision, data from wearables must be critically analyzed [4].

The acceleration in the adoption of digital health solutions was a direct response to global health crises, consolidating remote care as a pillar for maintaining medical assistance. Large-scale studies, such as systematic reviews, confirm that the patient experience with telehealth is largely positive, reaching levels of satisfaction that often equal or even surpass those of in-person interactions. This high approval is driven by remarkable convenience, savings in time and travel costs, and, especially, expanded access for geographically isolated populations or those with restricted mobility [5]. HR verification via video represents an expanding research field, valued for its convenient and contactless approach. The core methodology behind this technology, known as remote photoplethysmography (rPPG), is based on an optical principle: hemoglobin in the blood absorbs light distinctly. With each heartbeat, the volume of blood in the facial vessels subtly increases, altering the amount of light that is reflected back to the camera. Although these fluctuations in skin coloration are imperceptible to the naked eye, a conventional camera can register them. Image processing algorithms are then applied to isolate this pulsatile signal from the video data, extracting the heart rate. The accuracy of this technique, however, is highly dependent on factors such as the stability of ambient lighting, the individual's skin tone, and, especially, the absence of movement, which can corrupt the signal and hinder measurement [6].

A fundamental approach for contactless video analysis is Eulerian Video Magnification (EVM). This computational technique is designed to reveal phenomena that are imperceptible to the naked eye by selectively amplifying the minute temporal variations that occur in a sequence of images. Instead of tracking objects, the method analyzes the changes in color or intensity of each pixel over time. For HR measurement, EVM can isolate and amplify the frequency band corresponding to human heartbeats. This makes the blood flow under the skin visible as a subtle, pulsating color change. The main limitation of the technique is that it indiscriminately amplifies any signal—including noise—that falls within the selected frequency band. Consequently, the accuracy of the measurement can be compromised by patient movements or fluctuations in ambient lighting if these disturbances occur at a rhythm that overlaps with the cardiac signal [7].

An evolution in video magnification for motion analysis is the phase-based approach, which overcomes limitations found in previous methods, such as the Eulerian one. Unlike techniques that amplify the amplitude of color or brightness signals, this method uses the Hilbert Transform (or its 2D generalization) to isolate and amplify variations in the local phase of the video [8]. As phase is directly correlated with the position of a feature in the image, amplifying its changes over time translates into a more direct and cleaner magnification of motion. The result of this strategy is a superior ability to amplify larger magnitude movements without introducing the noise and visual artifacts that limited the previous technique. This allows for the visualization of motion phenomena on a broader scale and with greater fidelity, making the method more robust for applications in less controlled scenarios [9]. Another tool that can be introduced in the heart rate measurement scenario is MediaPipe, as a component for signal stabilization. Its ability to accurately detect and track facial landmarks in real-time allows for the definition of stable Regions of Interest (ROIs). It can provide up to 468 facial landmarks in the image, which represent the facial geometry in detail, accurately capturing the contours of the eyes, eyebrows, nose, lips, and the overall shape of the face.

The main contribution of this study is the presentation of a simplified and accessible pipeline for contactless HR monitoring. The key advancement lies in using the MediaPipe framework to automate the extraction of facial features, an approach that surpasses traditional methods in efficiency, which demand more intensive data preprocessing and rigorous capture setups. At the core of the methodology, the system analyzes the time series of the 468 three-dimensional facial landmarks provided in real-time by MediaPipe. This raw motion data is then subjected to a sequence of transformations: the Hilbert transform generates a complex analytic signal, PCA reduces its dimensionality, and finally, BSS precisely isolates the pulsatile signal of interest from other noise sources. The effectiveness of this process was rigorously validated, achieving results with high agreement when compared to a pulse oximeter used as a reference. The application potential of this technology is vast, as its optimized architecture shows promise for integration into embedded systems, such as those in smartphones and smart cameras, paving the way for the democratization of continuous health monitoring in everyday scenarios.

Theory

The measurement of HR from video motion analysis is based on the principle of remote ballistocardiography (rBCG). This phenomenon is a direct consequence of Newton's third law (action and reaction) applied to the hemodynamics of the human body. With each systolic contraction, the left ventricle ejects a volume of blood at high velocity into the ascending aorta; the "action" of this mass of blood being accelerated upwards generates an opposite "reaction": a subtle recoil impulse throughout the body. Although this recoil is minuscule, it propagates through the skeleton and, due to the biomechanics of the neck, manifests as a periodic and almost imperceptible oscillation of the head. It is this rhythmic movement that computer vision algorithms aim to detect [9].

The capture of the cardiac signal, in this method, does not originate from a global image analysis, but rather from the precise tracking of micro-variations in movement at specific facial points. First, the MediaPipe Face Mesh framework extracts the 3D coordinates of 468 landmarks over time, generating a rich time series of motion data [10]. This coordinate matrix then feeds into a sequence of three processing stages to isolate the heart rate: the Hilbert transform, PCA and BSS.

The methodology begins with the use of MediaPipe face mesh as a robust tool for facial feature extraction. Instead of analyzing raw video pixels, the system employs a sophisticated machine learning pipeline to detect and track 468 three-dimensional (3D) facial landmarks in real time. This process transforms the video into a high-density time series of coordinates, which provides a structured representation of the detailed movement of the facial surface. This data matrix serves as a rich and stable input for the subsequent signal processing steps, overcoming many of the instabilities associated with direct pixel analysis [10].

The analysis of subtle movements in video, such as those associated with physiological signals, often employs a combination of advanced signal processing techniques. The starting point is the application of the Hilbert Transform, which converts a real signal, such as the time series of a pixel's intensity, into its complex analytic counterpart. The main property of this analytic signal is that its imaginary part has a phase shift of exactly 90 degrees relative to the real part, a fundamental mathematical relationship for phase and envelope analysis [11] Figure 1 provides an overview of the proposed method.

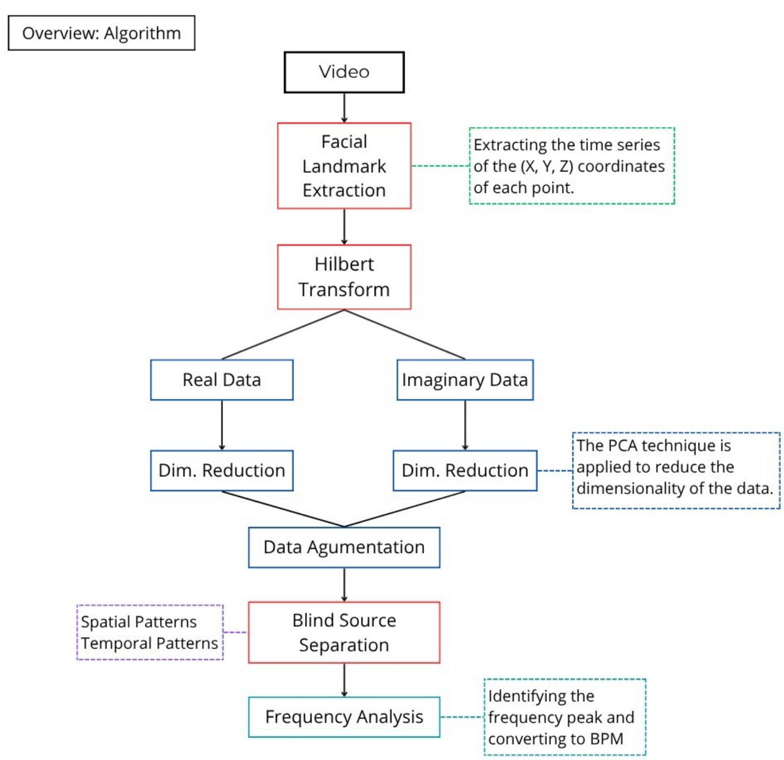

Fig. 1 Flowchart of the proposed approach

After applying the Hilbert Transform to the time series of facial landmarks, the resulting real and imaginary parts are concatenated, forming a single high-dimensional feature matrix, which will be called X. It is upon this matrix that PCA is applied with the goal of drastically reducing its complexity.

PCA is a linear transformation technique that projects data onto a new coordinate system such that the majority of the variance is captured in the first few coordinates, known as principal components. Mathematically, the process begins by calculating the covariance matrix C of the data, which has already been centered (by subtracting the mean from each column). The covariance matrix is given by:

$$C = \frac{1}{n} X^T X \tag{1}$$

where *n* is the number of samples (video frames) [12]. Next, the core of PCA consists of solving the eigenvalue and eigenvector problem for this matrix:

$$Cv = \lambda v \tag{2}$$

The eigenvectors v represent the principal components, which are the directions of maximum variance in the data, while the scalar eigenvalues λ indicate the magnitude of this variance. By ordering the eigenvectors according to their corresponding eigenvalues in descending order and selecting the first 10 components, as defined in the code, a reduced-dimensionality representation is obtained that preserves the most significant information of the facial movement. This compressed representation is fundamental to the pipeline's success, as it eliminates redundancies and noise. In doing so, PCA provides the subsequent BSS step with a cleaner and more focused dataset, optimizing the task of isolating the subtle physiological signal of the cardiac pulse. The eigenvectors v represent the principal components, which are the directions of maximum variance in the data, while the scalar eigenvalues λ indicate the magnitude of this variance. By ordering the eigenvectors according to their corresponding eigenvalues in descending order and selecting the first 10 components, as defined in the code, a reduced-dimensionality representation is obtained that preserves the most significant information of the facial movement. This compressed representation is fundamental to the pipeline's success, as it eliminates redundancies and noise. In doing so, PCA provides the subsequent BSS step with a cleaner and more focused dataset, optimizing the task of isolating the subtle physiological signal of the cardiac pulse.

After dimensionality reduction by PCA, the principal components are processed by the final stage of BSS. In the context of this work, BSS is used to separate the different facial movement patterns (captured by the PCA components) into their constituent sources, such as head movement, respiration, and, most importantly, the subtle pulsatile ballistocardiographic signal. The BSS problem is mathematically modeled by assuming that the observed signals *x(t)* (the PCA components) are a linear mixture of unknown independent sources *s(t)*:

$$x(t) = As(t) \tag{3}$$

The relationship involves A as the unknown mixing matrix and *s(t)* as the independent source signals. BSS aims to solve for both A and *s(t)* when only the observations *x(t)* are known. The process to recover the sources is completed by inverting the mixing matrix A after a reliable estimate for it has been found.

$$s(t) = A^{-1}x(t) \tag{4}$$

This work implements a specific BSS algorithm called complexity pursuit. The premise of this method is that the physiological signal of interest (the cardiac pulse) is more regular and, therefore, temporally more "predictable" (less complex) than other noises or random movements. The algorithm operates iteratively, adjusting the matrix W to minimize a cost function based on the unpredictability of each estimated source. The complexity function $F(w_i, x)$ to be minimized is:

$$F(w_i, x) = \frac{V(w_i, x)}{U(w_i, x)} \tag{5}$$

where $V(w_i, x)$ represents the long-term variance of the extracted signal, and $U(w_i, x)$ represents the short-term variance. The separation matrix W is found by maximizing the temporal predictability of the separated signals. The resulting signal sources correspond to the temporal patterns of each mode of movement in the scene.

Experimental results

The entire computational pipeline was developed in the Python language. For the experiments, video recordings were captured by smartphone cameras at a sampling frequency of 60 Hz, in everyday scenarios with uncontrolled lighting conditions. The method's accuracy was validated by comparing the results with measurements from a pulse oximeter, which served as the ground truth. The final analysis consists of calculating the power spectral density (PSD) of the isolated signal to identify the frequency with the highest energy. Subsequently, this peak in the frequency domain (f) is converted to the unit of beats per minute (BPM) through the following relationship:

$$\text{BPM} = 60 \times f. \tag{6}$$

During all trials, the participants were positioned in a frontal orientation to the camera. The comparative analysis revealed a minimal discrepancy between the heart rate estimated by the algorithm and the reference values provided by the oximeter, a result that robustly corroborates the effectiveness of the proposed methodology.

The initial validation of the methodology yielded encouraging results, prompting supplementary investigations to assess the system's robustness under varied conditions. To isolate the impact of eye movement artifacts, a trial was conducted in which the participant kept their eyes closed, making it possible to verify the influence of blinking. With the material used, it was possible to simultaneously test MediaPipe's facial landmark detection capability and the algorithm's sensitivity in capturing the pulsatile signal with spatial attenuation. Satisfactory results can be observed in Figures 2 and 3, and their comparisons with the ground truth in Table 1.

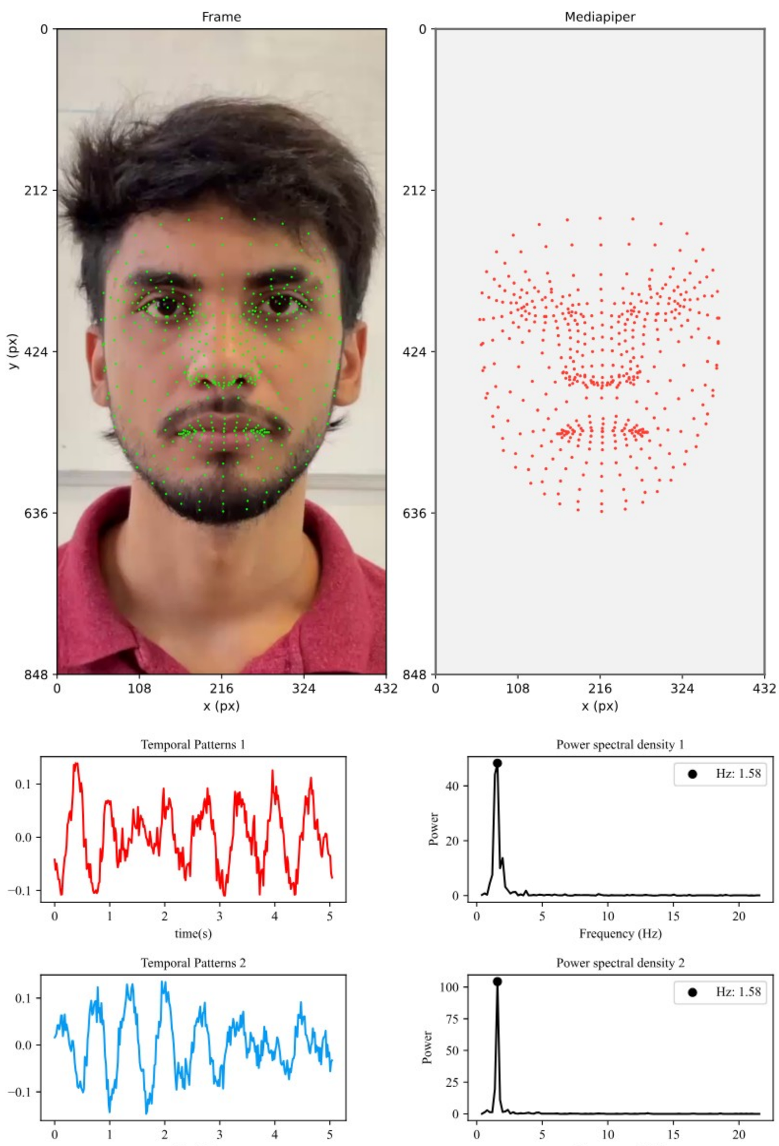

Fig. 2 Map of landmarks, modal coordinates and power spectral density from the test with

In all experiments, the error was insignificant, and variations in head angle did not significantly impact the results. Each setup successfully captured the subtle oscillations of the head caused by the pulsation of the aorta artery, and the heart rate was consistently detected with accuracy. MediaPipe facilitates the identification of points of interest, thereby reducing the amount of data. In this way, the processing did not utilize every pixel; a reshape was required to visualize the intensity of each pixel, and two pixels were added to form a complete 20x24 matrix.

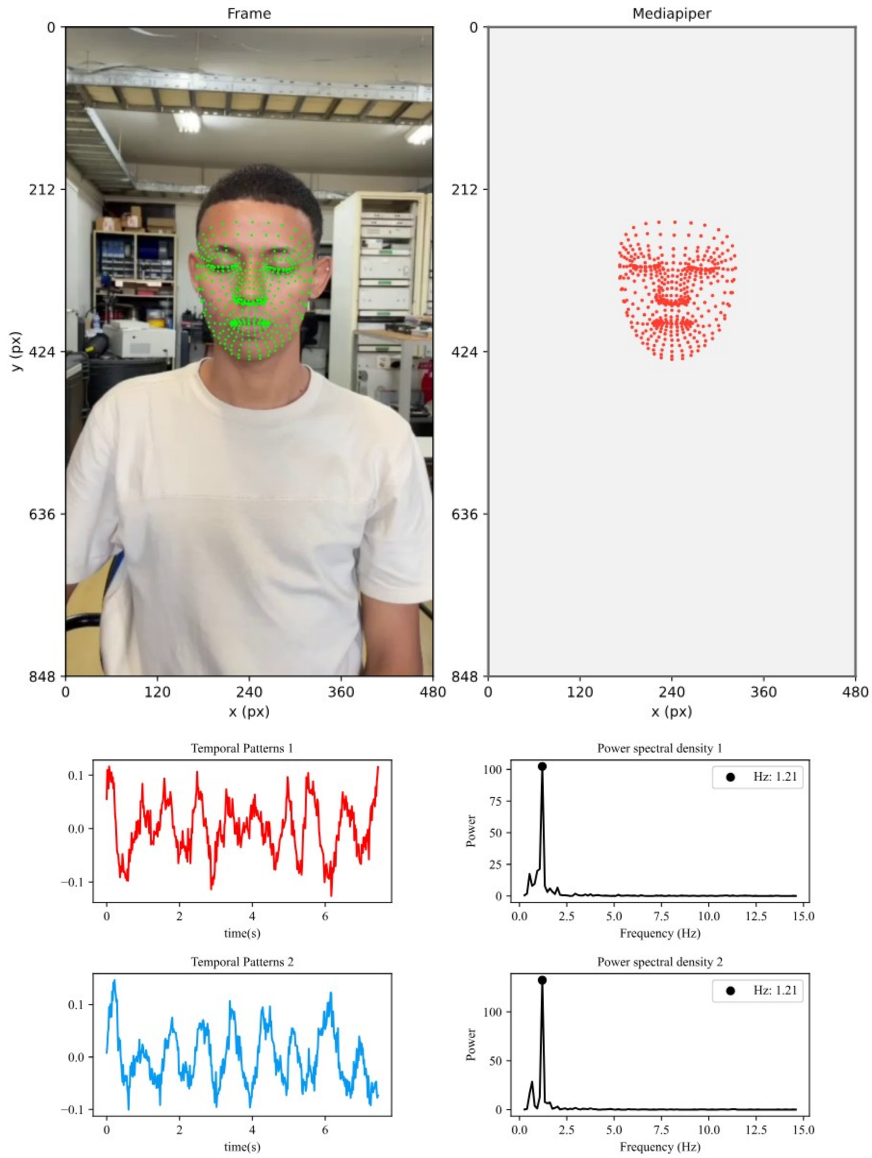

Fig. 3 Map of landmarks, modal coordinates and power spectral density from the test with

Table 1 Comparison of test results with oximeter readings in BPM

	Subject 1	**Subject 2**
Oximeter	100	73
Hilbert Magnification	94.8	72.6
Error (%)	5.2	0.4

Conclusion

This work presented and validated an innovative methodology for remote heart rate measurement, utilizing a computational pipeline that integrates facial landmark tracking via MediaPipe with advanced signal processing techniques, such as the Hilbert transform, Principal component analysis and blind source separation. The experimental results demonstrated the high effectiveness of the approach, achieving a remarkable agreement with measurements from a reference pulse oximeter.

The system's robustness was confirmed in tests under different conditions, including distance variations and eye occlusion, with the error margin remaining consistently minimal.

Therefore, it is concluded that the proposed approach is not only feasible, but also represents a significant advancement over previous methods, due to its simplicity and automation. The promising performance and efficient architecture pave the way for future optimizations and for the integration of this system into embedded devices, such as smartphones and smart cameras, aiming for the democratization of continuous health monitoring.

Acknowledgments It was supported by the Laboratory of Applied Electromagnetism (LEA) at the Federal University of Pará. Furthermore, the authors acknowledge that this study was financed in part by the Amazon Foundation to Support Studies and Research (FAPESPA), Grant agreement 050/2023. It was also financed by the São Paulo Research Foundation (FAPESP) with the support provided by grant 2022/10105-3, as well as the National Council for Scientific and Technological Development (CNPq).

References

1. Souza, A. et al. "Análise da sensibilidade e especificidade dos pontos de corte para frequência cardíaca de repouso como indicador de risco cardiometabólico em mulheres". *Arquivos Brasileiros de Cardiologia*, 115(4):670–677 (2020)
2. Domingues, W. et al. "A redução da frequência cardíaca após teste de esforço físico de intensidade progressiva está associada a melhor desempenho físico e menor risco cardiometabólico". *Arquivos Brasileiros de Cardiologia*, 117(3):517–525 (2021)
3. Silva, M. and Junior, J. "Aplicabilidade da variabilidade da frequência cardíaca como biomarcador do estresse: uma revisão integrativa". *Research, Society and Development*, 9(12):e3891211125 (2020)
4. Kaur, H. et al. "A comprehensive review on methodologies for heart rate monitoring and arrhythmia detection using photoplethysmography". *Sensors*, 23(18):7918 (2023)
5. Gudi, N. et al. "Patient satisfaction with telemedicine: a systematic review and meta-analysis". *Journal of Telemedicine and Telecare*, 29(5):315–326 (2023)
6. Villarroel, M. et al. "A survey on remote photoplethysmography". *ACM Computing Surveys*, 52(4):1–37 (2019)
7. Wu, H.Y. et al. "Eulerian video magnification for revealing subtle changes in the world". *ACM Transactions on Graphics*, 31(4):1–8 (2012)
8. Huang, N. et al. "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995 (1998)
9. Wadhwa, N. et al. "Phase-based video motion processing". *ACM Transactions on Graphics*, 32(4):1–10 (2013)
10. Kartynski, A. et al. "Mediapipe face mesh: Real-time 3d face surface geometry estimation on mobile devices". *arXiv preprint arXiv:2004.04865* (2020)
11. Torres, M., Rivera, M., and Gomez, J. "A review of time-frequency representation for the analysis of biomedical signals". *Signal, Image and Video Processing*, 15(8):1773–1781 (2021)
12. Bishop, C. *Pattern Recognition and Machine Learning*. Springer, New York (2006)