Comparative Analysis of Three Remote Photoplethysmography Models for Heart Rate Measurement During Simulated Blood Donation

A Study for Implementing a Heart Rate Measurement System in the AINAR Game

LAURA A. F. HEIJ

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN COGNITIVE SCIENCE
AND ARTIFICIAL INTELLIGENCE
AT THE SCHOOL OF HUMANITIES AND DIGITAL
SCIENCES OF TILBURG UNIVERSITY

STUDENT NUMBER

895724

COMMITTEE

dr. Elisabeth M.J. Huis in ‘t Veld

MSc. Fred Atilla

LOCATION

Tilburg University

School of Humanities and Digital Sciences

Department of Cognitive Science & Artificial Intelligence

Tilburg, The Netherlands

DATE

May 17th, 2023

WORD COUNT

7075

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to my supervisor Elisabeth for all the help and time she put into my bachelor’s thesis. Secondly, I want to acknowledge and thank Judita for all her assistance to overcome coding obstacles. I want to thank Welmoed and Lotte for taking experiments and thus providing data for my thesis. Lastly, I shortly want to thank my mom for rereading my thesis.

Comparative Analysis of Three Remote Photoplethysmography Models for Heart Rate Measurement During Simulated Blood Donation

A Study for Implementing a Heart Rate Measurement System in the AINAR Game

Laura A. F. Heij

Abstract

Numerous remote photoplethysmography (rPPG) models have been under development for a considerable period. However, most researchers tested their models on a dataset containing videos in a laboratory environment, with a lack of emotions from the participants. This thesis focused on determining the optimal performance among the three selected rPPG models (LGI, POS, and PCA) using our self-derived dataset. This analysis has been conducted to potentially implement the best-performing model into the game AINAR. The AINAR game helps people to conquer their needle fear, implying the model should be able to adequately measure heart rate fluctuations, as this is an indication of potential anxiety. To properly test if the model is sufficient enough for the AINAR application, we derived our own dataset. The dataset consisted of facial videos of participants undergoing a virtual blood donation, while their heart rate was tracked via a sensor-tracking device serving as ground truth. The videos were analyzed using an open-source Python-based framework pyVHR, containing a pipeline of rPPG models including the ones tested. The results of the contact heart rate tracker were compared with the three models. Results show that the LGI model is slightly better than the other two models (*MAE = 9.87*, *PCC = 0.08*). Although the results were not as sufficient as other research shows, the importance of meticulously obtaining one's own data is highlighted. These emphasized limitations led to several proposals for additional future work.

1 Data Source, Ethics, Code And Technology

For this study data is derived from participants of Tilburg University via SONA. Before conducting the experiment a consent form was signed by the participant and experimenter. The data was derived by recording the participant via sensory devices and a camera. Even though the face of the participant was recorded during the experiment, the data is anonymously implemented in this thesis. The owner of the data is my supervisor Elisabeth M.J. Huis in ‘t Veld and the blood bank Sanquin. The owner did give consent to use the data for this bachelor’s thesis. All the Figures used were created by the author. All the used libraries and packages are listed in the Method section under the subsection ‘Software’. Furthermore, the author used the code of PyVHR, accessible via GitHub (https://github.com/phuselab/pyVHR). The code written for this thesis can be found via the GitHub repository (https://github.com/Laura-0201/Bachelor-Thesis). To paraphrase my written text I used Thesaurus and ChatGTP (https://www.thesaurus.com/, https://openai.com/blog/chatgpt). To check the spelling, Grammarly (https://app.grammarly.com/) and the built-in spelling checker from Google Docs were used. To store the references of my sources Zotero (https://www.zotero.org/) was used.

2 Introduction

Needle fear is shared among the population and all ages, with more than 60% of children and on average 35% of adults experiencing it (McLenon & Rogers, 2019). However, blood donations and vaccinations are critical for saving lives. Therefore it would provide a significant advantage if needle fear could be reduced, so donations and vaccinations will increase. The mobile phone application AINAR (Artificial Intelligence for Needle Anxiety Reduction) is a game designed to help people conquer their needle fear through biofeedback. It assesses needle fear and the risk of fainting by analyzing the video stream of the front-facing camera. While playing the game, the AINAR app interprets the color change of the pigments on your face. The app will give the user real-time feedback while playing the game, ensuring the user gets notified of their well-being through the game before they notice it themselves, to prevent fainting or other discomforts. However, the current AINAR application does not consider heart rate as a factor, yet heart rate is a very informative aspect of physiological stress. A high value of beats per minute (bpm) as well as a low fluctuation of the heart rate (or in other words, low heart-rate variability) can indicate a form of anxiety (Kim et al., 2018). The most prevalent method to measure heart rate nowadays is via wearable technologies or electrodes. Not all people have access to these devices or can interpret the outcome correctly. However, facial video analysis algorithms have been developed to extract the heart rate via facial videos. This remote photoplethysmography (rPPG) models calculate the heart rate of a person without making physical contact through sensors. This technique measures the variance of colored light on the skin through an RGB camera (Rouast et al., 2018). The reflection of the skin changes due to the fluctuation of hemoglobin in the blood flow. This periodic change of hemoglobin can be captured by the camera due to color change and reflects the heart rate (*What Is RPPG?*, n.d.). However, these changes are very subtle, and movement or changes in light can cause noise, which leads to invalid predictions. The implementation of an rPPG-based heart-rate measurement technique may be an advantage of the already existing AINAR game. Until now, most research is done on participants sitting still under strict environmental conditions (Deng & Kumar, 2020). For a rPPG model to be relevant for the AINAR application, it should also be able to accurately track the fluctuations of the heart rate, as this might occur when a person is in a waiting room of a hospital or blood donation center.

Heart rate is an informative aspect of the human body, which tells a lot about the person's health and emotional experiences. However, as already mentioned nowadays heart rate is accurately measurable via wearables, which not everyone has access to (Ahmadi et al., 2022). Nonetheless, most people do have a phone with a front camera, which could record their faces at any time. Thus, it would be a significant advantage if a mobile device is able to measure heart rate using only video information, which would open up the possibility to measure heart rate to other researchers and clinicians. In this study, both video data and electrode-based heart-rate measurements are collected. This allows the comparison of the heart rate measured using rPPG from the video, with the heart rate measured with psychophysiological techniques. Furthermore, it is crucial to test the chosen models with data from participants undergoing an experience of something related to an environment in which needle insertion takes place. To test the model on relevant data, an experiment was set up. In this experiment participants were undergoing a fake blood donation, to create the dataset the models will be tested on. In short, this study provides a novel database collected using a paradigm similar to a real-life setting. Furthermore, this study will add to the literature assessing which type of rPPG method is most accurate on emotion-related data.

The goal of the thesis is to assess the reliability of several rPPG models when compared to heart rate measured with electrodes. To this aim, we use a Python-based application called PyVHR (Boccignone et al., 2022). PyVHR is a framework representing a multi-stage pipeline, capturing the process of extracting and analyzing the video and finishing by returning a heart rate in beats per minute. PyVHR gives the option to use nine different extraction techniques, of which three are chosen to compare. The rationale behind choosing these three techniques will be further elaborated on in the related work section.

Taking the goal into account, this thesis will address the following research questions.

> RQ1: *Using the pyVHR Python-based pipeline, which rPPG model (LGI, POS, PCA) best captures the heart rate from participants undergoing a virtual blood donation?*
>
> RQ2: *Does the best-performing model capture the fluctuations of the heart rate undergoing the virtual blood donation sufficiently?*

Concluded are several results reflecting the research questions. One of the main findings concluded that LGI performed best compared to the other two models, POS and PCA. However, the results exhibited minimal variation and were highly comparable. Secondly, the correlation between the true heart rate and the predicted one by the model was unexpectedly low. Indicating the chosen rPPG models are not satisfactory enough to implement in the AINAR game. However, there are some drawbacks of the measuring device used to set the ground truth of the heart rate, thus impacting the assessment of the model's performance as well.

3 Related work

Originally, electrodes are required for accurate results when measuring electrical heart activity using electrocardiographs (ECG). However, measurements that rely on physical contact are plagued by a pair of disadvantages (Yu et al., 2021). The topic of extracting heart rate from video files has been a rapidly developing technique and is called remote photoplethysmography (rPPG). It is based on photoplethysmography (PPG), which consists of a pulse oximetry device (van der Kooij & Naber, 2019). It uses infrared light to measure the volumetric variations in the arterial vascular network (Castaneda et al., 2018). Since a study revealed the feasibility of measuring PPG signals under natural light, interest in rPPG has grown (Verkruysse et al., 2008).

In the past, several signal-processing approaches have been developed, using multistep pre-processing steps (such as detrending), signal extraction, and post-processing steps (Lewandowska et al., 2011; Tsouri & Li, 2015). The primary constituents of an rPPG pipeline from video material to heart rate (HR) estimation contain consist of subject detection, face tracking, the detection of the region of interest (ROI), and using the pixels within the ROIs to extract the rPPG signal or to develop spatiotemporal maps on which neural networks were then used to estimate heart rate for a review, see (Cheng et al., 2021). To address the problems that can be caused by motion, several end-to-end deep learning-based methods were developed, these models take a raw video as input and produce a heart rate signal as the outcome. An overview of published models since 2018 can be found in the appendix named ‘Table 3: Overview of rPPG models since 2018’.

3.1 *Open source rPPG method: pyVHR*

For the purpose of this thesis, we chose to use PyVHR, a Python-based framework (Boccignone et al., 2022). PyVHR was selected because of the availability of user-friendly code, and the possibility to test nine different rPPG models within its pipeline. The nine traditional models are ICA (Poh et al., 2010), PCA (Lewandowska et al., 2011), GREEN (Verkruysse et al., 2008), CHROM (de Haan & Jeanne, 2013), POS (Wang et al., 2017), SSR (Wang et al., 2016), LGI (Pilz et al., 2018), PBV (De Haan & Van Leest, 2014), OMIT (Casado & López, 2023) and a deep learning-based method called MTTS-CAN (Liu et al., 2020).

The authors of pyVHR assessed the performance of the models on five datasets (Boccignone et al., 2020). Their results showed that PCA, SSR, POS, and CHROM are the best-performing models. However, a few of the datasets were of more interest than others. One of these datasets (Soleymani et al., 2012) consisted of videos of 30 participants varying in age, in an emotion induction experiment and included variation in viewpoints. On this dataset, models POS and PCA performed the best. Another dataset, LGI (Pilz et al., 2018) consisted of videos with more natural behavior, including head movements and variation in lighting. On that dataset, the PCA achieved a fairly good correlation, and the POS had the lowest Mean Absolute Error (MAE). Therefore, we decided to use the PCA and POS models. Additionally, the LGI model was used as this is one of the most recent ones developed and corresponds to a dataset in which the participants were in a non-laboratory environment. This may have benefits reflecting on our dataset.

3.2 *Description of the chosen methods*

3.2.1 *POS*

The POS method, built by Wang et al. (2017), has been widely used in recent years as a non-invasive and low-cost approach for monitoring heart rate using video-based imaging. The POS (plane-orthogonal-to-skin) approach is a signal decomposition model that, as the name asserts, generates a plane that is orthogonal to the skin tone within the RGB color space (Boccignone et al., 2022). First, the video file is preprocessed, and the input pixels of the video get normalized for changes in skin tone and lighting. Secondly, a skin segmentation algorithm is applied to determine the ROI. Next, a plane based on the orthogonal colors of the skin tone is created. The color signals of the ROI are projected on the created plane. The signals corresponding to the variations in color intensity along this plane are extracted. These signals representing the blood flow are used for extracting the pulse rate (Wang et al., 2017).
Several studies have shown that the POS technique has a high accuracy score, due to its capabilities such as robustness to changes in light, various skin tones, and different camera types. Additionally, the method is computationally efficient and can be easily implemented (Kossack et al., 2021; Pirnar et al., 2021). However, there are some limitations as well. It necessitates careful calibration of the projection plane characteristics and is susceptible to noise from external sources, motion artifacts, and variations in head posture. Wang et al. (2017) tested eight different rPPG models, evaluating skin tone, luminance, recovery, fitness, and overall performance. Based on the signal-to-noise (SNR) scores of the models, POS performed in most cases as best or second best. The overall best performance was achieved by POS (Wang et al., 2017).

3.2.2 *PCA*

PCA (principal component analysis) is an rPPG model which extracts the pulsatile component from the skin color signal, with the use of principal component analysis (Lewandowska et al., 2011). The fundamental idea of the PCA method decomposes the signals of the skin tones into various components. The component with the highest variability is used as the approximation for the changes in blood volume due to the cardiac cycle (Rios et al., 2021). The preprocessing is similar to those previously described for POS, containing skin segmentation and normalization of the video files. Next, principal component analysis is applied to the filtered signals to extract the component that corresponds to the highest variability in the data. This component is assumed to serve as a reliable approximation of the pulsatile component. Lastly, by applying a Fourier transform, or other spectrum analysis methods to calculate the extracted component's spectral peak, one may estimate the heart rate (Balakrishnan et al., 2013; Bogdan et al., 2015). Lewandowska et al. (2011) reported that the algorithm introduced demonstrates notable efficacy and user-friendliness for the routine monitoring of patients receiving home care. Lee et al. (2021) reported that PCA improved accuracy with respect to variations in illumination and the presence of motion artifacts. Furthermore, Boccignone et al. (2020) report PCA as one of the best models, statistically equal to POS, CHROM, and SSR.

3.2.3 *LGI*

Pilz et al. (2018) proposed the rPPG model LGI, trained to perform well under unconstrained natural settings (Boccignone et al., 2022). The LGI algorithm employs unsupervised learning to extract features from the blood flow signal that are invariant to the effect of differentiable local transformations (Ou et al., 2022).

This is achieved by re-arranging the signal into a more concentrated distribution. When compared with five other algorithms, including POS, the LGI algorithm demonstrates significant improvement in certain scenarios, particularly those involving head rotation and speech (Deng & Kumar, 2020). The study of Deng & Kumar. (2020) compared eight different rPPG models on the LGI-PGGI dataset made by (Pilz et al., 2018), this dataset contained six participants all recorded during rest, talking, head rotation, and gym activities. For the gym, talking and head rotation sessions LGI did outperform the other tested models. However, LGI did perform slightly worse in the resting stage. Found was that under all conditions LGI, POS, and one other model performed best, with a slight advantage for LGI. Despite having a better statistical performance, the LGI approach exhibits an estimation bias of about 4 bpm. This also explains why, in the resting case, the LGI approach performs a little worse (Pilz et al., 2018).

3.3 *Evaluation of rPPG algorithms*

To evaluate the accuracy of the rPPG results, they are compared to the heart-beat collected using a Shimmer. Common evaluation techniques are SNR, Bland-Altman Plot, Pearson's Correlation Coefficient, Error rate, and ANOVA (Deng & Kumar, 2020). The most relevant techniques for this research and also mentioned by Boccignone et al. (2022) are Pearson's Correlation Coefficient and the Error rate measured by the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). PCC measures the linear correlation between two variables, which in this study corresponds with the correct rising or dropping of the heart rate. The PCC represents the covariance from both variables divided by the product of their standard deviations. It returns a value between -1 and 1, representing the linear correlation. This thesis aims for a value as proximate as feasible to 1, this suggests the true heart rate has the same direction as the predicted one. The second evaluation matrix, RMSE, computes the square root of the mean of the squared differences between two values. The RMSE measures the discrepancy between the ground truth heart rate and the heart rate computed by the model (Deng & Kumar, 2020). Lastly, the MAE is calculated, which measures the average size of mistakes in a set of forecasts. In fact, the arithmetic mean of the absolute differences between the expected and observed values, with equal weight given to each individual difference (JJ, 2016).

3.4 *Closing the research gap*

Over the past decade, the technology on rPPG has been significantly improved, due to more datasets, camera advantages, and computational power. Nevertheless, there are still drawbacks and challenges within the field. Even though some researchers take motion, light changes, and other noise into account, most of them are conducted in a laboratory environment (Sabokrou et al., 2021). Moreover, the majority of studies focus on to what extent the average 'beat-to-beat' over a long period of time is the same using rPPG or other psychophysiological methods. However, it is still not apparent if rPPG methods are adequate for capturing instantaneous HR changes (Deng & Kumar, 2020). Reflecting on my research, capturing fluctuating heart rates is likely to be a crucial factor for data from the virtual blood donation experiment. To partly close this research gap, we will also assess whether the signal of both methods corresponds in time.

4 Method

4.1 *Software*

The measurements of the Shimmer were recorded and stored by the software of the Shimmer called Concensys (*Consensys Software*, n.d.). For the preprocessing, processing, and analysis part of the video and Shimmer files, several codes were written in Python (Drake & van Rossum, 2009). To format the data properly the libraries Numpy (van der Walt et al., 2011), Pandas (McKinney, 2010), and datetime (*Datetime*, n.d.) were used. The videos were preprocessed using the libraries CV2 (Bradski, 2000) and Moviepy (Zulko, 2017). The transcription of Shimmer files was performed utilizing the libraries Heartpy (van Gent et al., 2018) and Neurokit2 (Makowski et al., 2021). To analyze and visualize the results the following packages were used: Matplotlib (Hunter, 2007), Scipy (Virtanen et al., 2020), Sklearn (Pedregosa et al., 2011), and Math (Van Rossum, 2020).

4.2 *Data collection*

4.2.1 *Sample and procedure*

The sample consisted of N = 53 participants (*M = 20.96*, *SD = 2.49*), of which N = 17 were men (*Mage = 21.41*, *SDage = 3.34*) and N = 36 were women (*Mage = 20.76*, *SDage = 1.95*). The participants were recruited at Tilburg University, through SONA. After providing informed consent, the participant completed a questionnaire with items regarding personal characteristics and needle fear. Then, the baseline measure of heart rate was collected, via the Shimmer device. In the meanwhile, the participant was scrolling through a phone given by the experimenter, in which they had 4 apps (BBC News, BBC Sport, Science News, Flipboard) they could vary between. This phase lasted approximately 4 to 5 minutes. It should be noted that the data is collected by four different experimenters, which might have influenced consistency of the data.

4.2.2 *The virtual blood donation experiment*

Then, the participant underwent a virtual blood donation experiment. The aim of this experiment is to induce the so-called rubber arm illusion, a phenomenon that elicits comparable reactions to events that happen to the arm as if it was the participants real arm. The method of this experiment is originally from the study of Trost et al. (2017) and was also implemented by the study by Rudokaite et al. (2022). First, the participant was asked to place their arm behind the laptop screen. The participants visual field was limited to the screen, where a virtual arm was presented. The experimenter had access to twelve different video's representing different skin colors, whether it was a female or male, and the left or right arm. To create the illusion, the experimenter copied the movements shown on the screen. These movements consisted of a brush, striking the arm and fingers of the participant, followed by disinfecting the needle insertion place with a wet cotton ball, and finally the needle insertion itself. However, water was used instead of disinfecting fluids, and a caliper was used to represent the needle. The participants are equally and randomly assigned to be in a synchronous (meaning the experimenter copied the movements on the screen exactly) or asynchronous condition (in which the movements on the real arm are delayed to disrupt the illusion). Lastly, we assessed whether the illusion was induced using the questionnaire Rubber-hand illusion (part 2).

4.2.3 *Measurement devices*

The faces of the participants were filmed during baseline and the experiment phases, using a digital camera placed on a tripod. This camera has an accuracy rate of 25 frames per second. Additionally, the phone during the baseline phase, and the laptop during the experimental phase were also recording the face. However, these videos will not be used in this research. Heart rate and skin conductance are measured using a Shimmer3 GSR+ Unit, placed around the wrist and attached to two of the participants fingers and one earlobe. These three contact points with the body monitor the skin conductance, of which the results get presented in a photoplethysmograph (PPG) (*Shimmer3 GSR+ - IMotions*, n.d.). The Shimmer3 GSR+ Unit has a sample rate of 128 Hz, meaning 128 measurements per second.

The relevant data conducted for one participant consisted of one facial video during the fake blood donation and the corresponding Shimmer file. The relevant data from the Shimmer file consisted of the Unix time as well as the associated PPG signal, from which the heart rate in bpm can be derived.

4.3 *Preprocessing*

4.3.1 *Preprocessing Video files*

The original videos were cropped to subtract the relevant part of the video containing only the rubber arm illusion. This was done with the Python library Moviepy (Zulko, 2017). Moviepy is a Python library designed for video editing. The function sub-clip was used to cut the beginning and end of the video. The start of the cropped video aligns with the commencement of the blood donation, and the end of the cropped video corresponds to the end of the blood donation. This choice was taken since it is clear that all participants data accurately reflects their shared experience. This resulted in videos from 2 minutes and 2 seconds up to 2 minutes and 59 seconds. After cropping the videos, the length of the participants differs due to the different videos chosen by the experimenter. By cutting the video irrelevant heart rates are disclosed from the dataset.

4.3.2 *Preprocessing Shimmer files*

The raw Shimmer datafiles contained Unix time and the corresponding PPG signal over the baseline measurements, fake blood donation, and the short time period in between, resulting in a file length of about 12 minutes. First of all, the Unix time was translated to standard time. With the use of the metadata from the video, the exact starting time from the video was determined. This time should correspond with the starting time of the Shimmer file. Therefore, the Shimmer file was cropped, containing only the time and PPG signal associated with the video. This process was carried out manually, due to the limited number of files and to improve accuracy.

4.3.3 *From PPG to bpm*

To check whether the generated PPG signals from the Shimmer file are adequate to generate a corresponding bpm the Python library Heartpy (van Gent et al., 2018) was used. This library is designed to transform noisy PPG data to bpm. Gent et al. (2018) claim that Heartpy is a robust well-performing algorithm. The algorithm contained cleaning the PPG signal to afterward predict the bpm. Nevertheless, even after cleaning the PPG, the signal was either unreadable for

Heartpy, or led to unbiological high bpm. As these videos do have no reliable ground truth, these participants were eliminated from the dataset.

4.3.4 *Data selection and outliers*

A total of N = 7 participants were excluded from analyses because the acquired data from the Shimmer files were of too low quality, or they resulted in heart rates that are biologically impossible. This was probably due to a technical problem with one of the Shimmers. Also, some files were missing due to equipment failures or mistakes by the experimenter. Furthermore, some video files could not be used, because the face of the participant was not properly filmed or the face of the experimenter was visible.

The preprocessing of the data resulted in a dataset of 7 participants, which were classified to have valid and reliable data for both the Shimmer file and the video. These participants were further analyzed in the processing part.

4.4 *Data Processing*

4.4.1 *Processing the videos with pyVHR*

The PyVHR model is an open-source model consist, which is freely accessible on Github (Grossi, 2023). The model is composed of a pipeline that accepts the cropped video as input and generates the corresponding bpm for each time window as output. There are several parameters that can be modified to attain varied performance outcomes. The most important parameter is ‘method’, corresponding to the rPPG method used out of the nine options mentioned in the related work section. Second, the ROI method and approach can be chosen, the options ‘convexhull’ and ‘patches’ were chosen respectively. The parameter ‘bpm_type’ was left with the default setting ‘welch’. ‘Post_filt’ and ‘Verb’ were set to ‘True’, while ‘pre_fit’ and ‘cuda’ were set to ‘False’. Cuda must be set to ‘False’ as this is exclusively accessible on a GPU device. It should be noted that implemented code of the PyVHR is slightly different from the one represented on the GitHub page, this was due to technical issues. This will be further discussed in the discussion section. Lastly, all 7 videos were run through the model for the 3 chosen rPPG models LGI, POS, and PCA, resulting in 21 files of estimated heartbeats. These files constituted an array of bpm approximately every 25.8 frames, which results in a bpm for a little bit more than a second as the video frame rate was 25 frames per second.

4.4.2 *Extracting bpm with Neurokit2*

To process the PPG signal into bpm the Python library Neurokit2 was used (Makowski et al., 2021). Neurokit2 is a Python package providing tools to reduce noise and improve signal processing. First, the raw signal of the cropped Shimmer file is cleaned by the function ppg_clean, which included preparing a raw PPG signal for peak detection. This is followed by the processes of finding peaks, the corresponding function is ppg_process. The function returns a data frame with the same length as the original PPG file, containing the bpm for each time point. For both two mentioned functions the default method Elgidi was selected. This method was chosen as their research demonstrates the highest accuracy (Elgendi et al., 2013).

4.5 *Evaluation criteria*

The process resulted in a ground truth of a participant and three corresponding outputs of the pyVHR model. However, the Neurokit2 library returns a bpm for each input PPG value, resulting in different lengths for the Shimmer and video bpm files. Nevertheless, concluded was that the length of both the video and Shimmer data are equal, as they are cropped at the same time. To properly compare the video results with the ground truth, the time window in which the heart rate gets calculated should be the same. Therefore, the amount of bpm of the Shimmer file gets reduced to the same amount as the pyVHR. This is done by taking the mean over an epoch of the same time window as the pyVHR. Resulting in the corresponding bpm in time for both the ground truth and the estimated one.

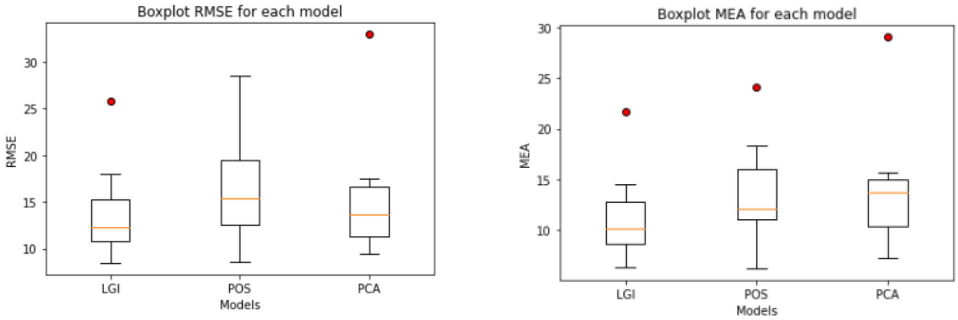
First of all, to compute the accuracy of the POS, PCA, and LGI models as compared to the bpm measured by the Shimmer, per participant, the Root Mean Standard Error (RMSE) was computed, as this gives a higher penalty to outliers. Additionally, to compare the linear trend over time, the Pearson Correlation Coefficient (PCC) was computed. Furthermore, to compare with other research, the Mean Absolute Error (MAE) was also computed. These techniques were already mentioned in the related work section. Moreover, these statistical analyzes were chosen based on the recommendations in the article of Boccignone et al. (2022). The mean and the Standard Deviation (SD) of all participants were taken to conclude the performance of the model. This process was done for all three individual models. Resulting in three mean RMSE, MAE, and PCCs with corresponding SD.

5 Results

5.1 *Outliers*

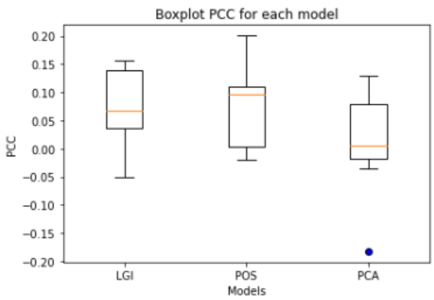
Before comparing the results of the three models, outliers were assessed. For both the RMSE and MAE Figure 1 shows a consistent red-colored outlier, which corresponds with participant 16.

Figure 1
Boxplots and outliers on RMSE and MAE



Additionally, the PPC boxplot also contains a significant outlier represented by a blue dot in Figure 2, corresponding to participant 11. The next section will look in more depth into participant 16 and conclude whether or not to exclude it from further analysis.

Figure 2
Boxplot and outlier on PCC



5.2 *Results per participant*

In this section, participants and their model performance will be highlighted to provide a clearer picture of the findings. Table 1 shows the performance of the models on each individual participant. The lowest RMSE is 8.48, achieved by LGI on participant 5. An RMSE of 8.48 means that the difference between the predicted values and actual values is on average off by 8.48 bpm. The lowest MAE found is 6.17 from POS, representing the absolute differences. The worst RMSE of 32.94 and MAE of 29.08 was found by the PCA model for participant 16. All heartbeat estimations will be off by approximately 8 to 33 bpm.

Table 1

*Performance metrics for each rPPG model and participant. * p-value < 0.05, ** p-value < 0.01*

	MAE			RMSE			PCC		
Pt.	PCA	LGI	POS	PCA	LGI	POS	PCA	LGI	POS
5	**7.23**	**6.25**	**6.17**	**9.52**	**8.48**	**8.63**	**0.06**	0.07	-0.01
6	12.0	9.320	8.94	15.69	12.34	11.69	-0.04	**0.16***	0.11
11	10.86	10.13	12.29	13.69	12.72	15.43	-0.18*	0.05	0.11
16	**29.08**	**21.66**	**24.18**	**32.94**	**25.77**	**28.50**	0.01	0.02	0.02
39	11.66	11.17	13.64	12.24	12.18	17.16	0.13*	**0.28****	0.20*
44	8.91	7.86	11.18	10.49	9.47	13.31	-0.01	0.15	0.10
52	14.36	14.46	18.37	17.50	17.95	21.86	0.01	-0.05	-0.02

As is apparent from the data provided in Table 1, participant 16 performed poorly for all models. This could be due to the heart rates measured by the Shimmer being overall much lower than those extracted from the videos (and lower than what is normally expected), which raises the question of whether the ground truth in this case is biologically plausible. Therefore, given that participant 16 was also an outlier showed in the boxplots, the results are excluded from the follow-up analyses.

Furthermore, the Pearson correlation coefficients are all close to zero, which means that there is (almost) no association between the bpm measured with the videos and those measured by the Shimmer. This suggests that the trend of rising and dropping heart rate is very poorly captured. The results are graphically represented in a Figure 3 and 4, illustrating the data from the models as well as the Shimmer data for participant 5 and 16. Figure 3 corresponding to participant 5 shows that the range of the Shimmer heart beats correspond to the range predicted by the models. However, as Figure 4 shows, the Shimmer results lay far below the heart beats estimated by the model. The visualizations of the other participants can be found in the appendix.

Figure 3

Overall extracted heart rate using LGI, POS, and PCA models compared with the Shimmer heart rate for participant 5.

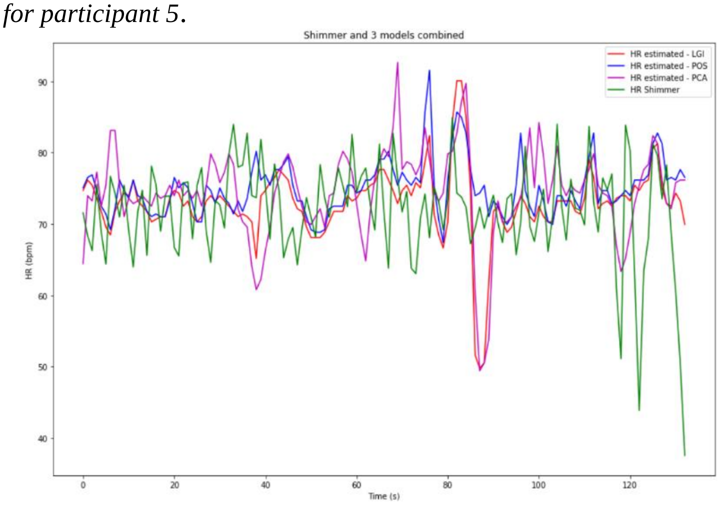
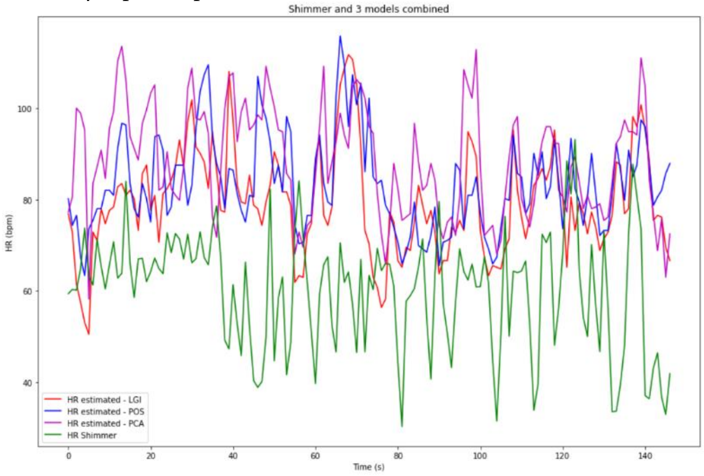


Figure 4

Overall extracted heart rate using LGI, POS, and PCA models compared with the Shimmer heart rate for participant 16.



In addition to the standout best and worst performers, there are noteworthy results among other participants that deserve mention. First of all, following participant 16, the performance of participant 52 is consistently the lowest across all models. Focusing on the PCC, most of the results are not statistically significant. However, it can be seen that participant 39 has a significant correlation for all 3 models. The LGI model even has a significance of a p-value < 0.01. Another noteworthy outcome is the PCC of the PCA model from participant 11. As this shows significant results, yet is a negative correlation, implying the direction of the heart rate of the model does the opposite of the true heart rate. In the discussion, these remarkable findings will be further elaborated on.

5.3 *Best-performing models*

To conclude this section, the three models are compared based on their RMSE, MAE, and PCC (see Table 2). First off all, it can be seen that the differences between all models are not exceptionally significant. However, overall best performing for RMSE, and MAE is the model of LGI. POS and LGI can be classified as equally good capturing the trend of the heart rate evaluated by PCC. However, POS performed the worst based on RMSE and MAE. PCA is second best at predicting the heartbeat, yet performs very poorly when capturing the trend of the heart rate.

Table 2

Performance of the models excluding participant 16 MAE (Mean Absolute Error), RSME (Root Mean Square Error), and PCC (Pearson's correlation).

	RMSE		MAE		PCC	
	Mean	Std	Mean	Std	Mean	Std
LGI	**12.19**	3.31	**9.87**	2.84	**0.08**	0.08
POS	14.68	4.60	11.77	4.17	**0.08**	0.08
PCA	13.19	3.06	10.84	2.50	-0.01	0.10

6 Discussion

6.1 *Reflecting the resulted on the RQs*

The first RQ was: ‘Using the pyVHR Python-based pipeline, which rPPG model (LGI, POS, PCA) best captures the heart rate from participants undergoing a virtual blood donation?’ The best-performing model was LGI. However, this model does not perform significantly better than POS and PCA. The RMSE and MAE were slightly higher than expected, while the PCC is way worse than expected. Boccignone et al. (2022) tested all nine models on different datasets and therefore has varying results. However, the highest MAE they reported is 21.07 for POS on the LGI-PPGI-gym (Pilz et al., 2018), and the lowest value of 0.9 for POS on the dataset PURE-slow-trans (Stricker et al., 2014). The median MAE for POS, PCA, and LGI respectively is 2.07, 3.01, and 3.27. This suggests that our results are below the abilities of the pyVHR model. However, the PURE-steady data is maybe not a representative comparison, as our data consisted of moving, sometimes talking participants, not ordered to sit completely still. As the LGI dataset contains moving and talking people, this is maybe a more represented way to compare our results.

The second RQ was: Does the best-performing model capture the fluctuations of the heart rate undergoing the virtual blood donation sufficiently? The best-performing model is LGI, with a corresponding best PCC of 0.08. This result is not statistically significant and therefore we can conclude that the model based on our results does not capture the fluctuations of the heart rate sufficiently. Comparing the results from Boccignone et al. (2022), the LGI median PCC of 0.41. The best-tested dataset UBFC1 with a value of 0.68 and the worst of -0.04, followed by the second worst with a value of 0.24. Our work exhibited significantly lower performance in terms of correlations. Another thing to note is that in Boccignone et al. (2022) research, both POS and PCA have a higher median PCC, with 0.73 and 0.60 respectively.

6.2 *Highlighting participants*

As discussed in the result section there were notable and exceptional outcomes. This subsection will delve into the factors contributing to the enhanced accuracy of participant 5's video. Additionally, it will explore the reasons behind participant 39 achieving notably high correlation levels. On the other hand, discuss why participant 16 and 52 perform so poorly. While inspecting the individual videos, some remarkable things did arise. First of all, the face of participant 5 was filmed the most frontal of all, other participants were looking more down- or sidewards towards the laptop screen. Therefore, it may be important that the best performance is related to the angle at which the face is filmed. The frame of participant 16 was more zoomed out than the other videos, this could have an influence on the performance. The recording of participant 52 is a little bit blurry, which could have influenced the performance of the model. However, the video of participant 39 was equally blurry but performed best based on the PPC score, which suggests sharpness does not have a huge influence. To conclude, the angle at which the face has been filmed does seem to have the greatest influence on the performance of the pyVHR model.

6.3 *Limitations of the study*

Even though research suggests that the Shimmer hardware is a well-performing device (Ràfols-de-Urquía et al., 2019), in this study we had to discard most of the participants due to the poor quality of the Shimmer data. This means unfortunately that the results of the rPPG models are very hard to quantify, as this data served as the ground truth. We do need to highlight that we did not have access to the pro version which allows an on-site check of the recordings. Furthermore, we did not have access to the ECG electrodes (*Consensys ECG Development Kits*, n.d.) but had to use the GSR module (*Shimmer3 GSR+ Unit*, n.d.) which captures bpm using PPG. Future researchers are highly advised to either use the pro version or to use another way of visualizing the measurements taken during the experiment in real-time, to check and fix any errors, and to possibly use ECG electrodes.

Consequently, this study describes the results of only 7 out of 53 participants, which influences the reliability of the overall results. Due to the involvement of different experimenters in administering the experiment to the participants, the results obtained varied, potentially introducing biases into the collected data. Furthermore, the original pipeline provided by the researchers of pyVHR was not compatible with the hard- and software used. This led to manually installing the pipeline, resulting in some arguments abilities that become unavailable. One of these features which may have influenced the output is the window size in seconds. Boccignone et al., 2022 obtained the best PCC results when setting the window size between 7 and 10. While for us this was set to the default setting of 5.

6.4 *Clinical implications*

The higher-order purpose of the study was to reason whether AINAR could be implemented with an rPPG model. Purely based on our own result it would not be suggested to implement one of the models in the AINAR game, as the error rate is too high. However, with all the limitations encountered, such as the number of participants, no optimal measurement materials, and the quality of the data, it may not be fair to judge based on this result. Furthermore, Boccignone et al. (2020) showed that there are rPPG models which do have sufficient results on different datasets. To properly conclude whether AINAR should implement an rPPG model and which one, the limitation first should be solved.

6.5 *Future work*

Based on the previously discussed limitation and results, there are some suggestions for future work. First of all, the models could not properly be evaluated, because the ground truth formed by Shimmer could not be fully trusted. Would this experiment be done once again a more accurate tracking device should be used, to give a valid report about the results of the model. Additionally, it is recommended to assign a single experimenter to conduct the experiment in order to minimize data variation. There was also concluded that persons who were filmed frontally gave the best results. The experiment contained three filming devices, the camera, phone, and laptop. As most participants looked at the screen during the experiment, the recording of the laptop may be more frontal than the video used from the camera on the triplot. May the original data set be used, it is suggested to also take into account the camera recordings of the laptop. A possible downside to this is the quality of these recordings. Furthermore, this research only looked at three rPPG models, while future research could also investigate the other six models proposed in the pyVHR pipeline. Additionally, recent research showed an effective rPPG method based on deep learning, which could also be an interesting application for our exclusive dataset (Sun et al., 2023).

7 Conclusion

The aim of this study was to find the best rPPG model to implement into the game AINAR. As a result, the research goal was established; investigating which of the three chosen rPPG models (POS, PCA and LGI) performed best using the pyVHR framework. Performance was evaluated based on the accuracy of the bpm and the correlation of the fluctuation of the heart rate. The results showed that the LGI model slightly outperformed the other two tested models POS and PCA, both at accuracy and correlation. However, related work showed better performance for all three models on various datasets (Boccignone et al., 2022). Moreover, there were several limitations such as the number of participants and valid heart rate tracking devices. This sparked a discussion regarding the feasibility of effectively evaluating the models. Additionally, this research added exclusive data to the existing field, as the derived data consisted of participants showing real emotions while undergoing a fake blood donation, such as fear and anxiety. To make informed decisions about the most suitable rPPG model and its implementation into AINAR, it is crucial for future research to consider the mentioned limitations and thoroughly assess the performance differences among the models.

References

Ahmadi, N., Al Farisyi, M. S., Prihatmoko, M. D., Hyanda, M. H., Muhaimin, H., Mulyawan, R., Charlton, P. H., & Adiono, T. (2022). Development and Evaluation of a Contactless Heart Rate Measurement Device Based on rPPG. *2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 1–4. https://doi.org/10.1109/ICECS202256217.2022.9971006

Balakrishnan, G., Durand, F., & Guttag, J. (2013). *Detecting Pulse from Head Motions in Video*. 3430–3437. https://openaccess.thecvf.com/content_cvpr_2013/html/Balakrishnan_Detecting_Pulse_from_2013_CVPR_paper.html

Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., & Lanzarotti, R. (2020). An Open Framework for Remote-PPG Methods and Their Assessment. *IEEE Access*, *8*, 216083–216103. https://doi.org/10.1109/ACCESS.2020.3040936

Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., Lanzarotti, R., & Mortara, E. (2022). pyVHR: A Python framework for remote photoplethysmography. *PeerJ Computer Science*, *8*, e929. https://doi.org/10.7717/peerj-cs.929

Bogdan, G., Radu, V., Octavian, F., Alin, B., Constantin, M., & Cristian, C. (2015). Remote assessment of heart rate by skin color processing. *2015 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 112–116. https://doi.org/10.1109/BlackSeaCom.2015.7185097

Bradski, G. (2000). *The OpenCV Library*.

Casado, C. Á., & López, M. B. (2023). *Face2PPG: An unsupervised pipeline for blood volume pulse extraction from faces* (arXiv:2202.04101). arXiv. https://doi.org/10.48550/arXiv.2202.04101

Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C., & Nazeran, H. (2018). A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors & Bioelectronics*, *4*(4), 195–202. https://doi.org/10.15406/ijbsbe.2018.04.00125

Cheng, C.-H., Wong, K.-L., Chin, J.-W., Chan, T.-T., & So, R. H. Y. (2021). Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda. *Sensors*, *21*(18), 6296. https://doi.org/10.3390/s21186296

Consensys ECG Development Kits. (n.d.). Shimmer Wearable Sensor Technology. Retrieved May 11, 2023, from https://shimmersensing.com/product/consensys-ecg-development-kits/

Consensys Software. (n.d.). Shimmer Wearable Sensor Technology. Retrieved May 11, 2023, from https://shimmersensing.com/product/consensyspro-software/

Datetime. (n.d.). Python Documentation. Retrieved May 11, 2023, from https://docs.python.org/3/library/datetime.html

de Haan, G., & Jeanne, V. (2013). Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Transactions on Biomedical Engineering*, *60*(10), 2878–2886. https://doi.org/10.1109/TBME.2013.2266196

De Haan, G., & Van Leest, A. (2014). Improved motion robustness of remote-PPG by using the blood volume pulse signature. *Physiological Measurement*, *35*(9), 1913–1926. https://doi.org/10.1088/0967-3334/35/9/1913

Deng, Y., & Kumar, A. (2020). Standoff heart rate estimation from video: A review. *Mobile Multimedia/Image Processing, Security, and Applications 2020*, *11399*, 16–29. https://doi.org/10.1117/12.2560683

Drake, F. L., & van Rossum, G. (2009). *Python 3 Reference Manual*. CreateSpace.

Elgendi, M., Norton, I., Brearley, M., Abbott, D., & Schuurmans, D. (2013). Systolic Peak Detection in Acceleration Photoplethysmograms Measured from Emergency Responders in Tropical Conditions. *PLOS ONE*, *8*(10), e76585. https://doi.org/10.1371/journal.pone.0076585

Grossi, G. (2023). *Github/PyVHR*. https://github.com/phuselab/pyVHR

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

JJ. (2016, March 23). MAE and RMSE — Which Metric is Better? *Human in a Machine World*. https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, *15*(3), 235–245. https://doi.org/10.30773/pi.2017.08.17

Kossack, B., Wisotzky, E., Hilsmann, A., & Eisert, P. (2021). *Automatic Region-Based Heart Rate Measurement Using Remote Photoplethysmography*. 2755–2759. https://openaccess.thecvf.com/content/ICCV2021W/V4V/html/Kossack_Automatic_Region-Based_Heart_Rate_Measurement_Using_Remote_Photoplethysmography_ICCVW_2021_paper.html

Lee, H., Cho, A., & Whang, M. (2021). Fusion Method to Estimate Heart Rate from Facial Videos Based on RPPG and RBCG. *Sensors*, *21*(20), 6764. https://doi.org/10.3390/s21206764

Lewandowska, M., Rumiński, J., Kocejko, T., & Nowak, J. (2011). Measuring pulse rate with a webcam—A non-contact method for evaluating cardiac activity. *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 405–410.

Liu, X., Fromm, J., Patel, S., & McDuff, D. (2020). Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement. *Advances in Neural Information Processing Systems*, *33*, 19400–19411. https://proceedings.neurips.cc/paper/2020/hash/e1228be46de6a0234ac22ded31417bc7-Abstract.html

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

McKinney, K. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 51–56.

McLenon, J., & Rogers, M. A. M. (2019). The fear of needles: A systematic review and meta-analysis. *Journal of Advanced Nursing*, *75*(1), 30–42. https://doi.org/10.1111/jan.13818

Ou, W., Chen, L., Han, J., Xiong, J., Zeng, W., & Gou, J. (2022). *Non-Contact Heart Rate Measurement from Face Video Based on Group Sparse Representation* (SSRN Scholarly Paper No. 4291045). https://doi.org/10.2139/ssrn.4291045

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., & Grisel, O. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12(Oct)*, 2825–2830.

Pilz, C. S., Zaunseder, S., Krajewski, J., & Blazek, V. (2018). *Local Group Invariance for Heart Rate Estimation From Face Videos in the Wild*. 1254–1262. https://openaccess.thecvf.com/content_cvpr_2018_workshops/w27/html/Pilz_Local_Group_Invariance_CVPR_2018_paper.html

Pirnar, Ž., Finžgar, M., & Podržaj, P. (2021). Performance Evaluation of rPPG Approaches with and without the Region-of-Interest Localization Step. *Applied Sciences*, *11*(8), 3467. https://doi.org/10.3390/app11083467

Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, *18*(10), 10762. https://doi.org/10.1364/OE.18.010762

Ràfols-de-Urquía, M., Estrada, L., Estévez-Piorno, J., Sarlabous, L., Jané, R., & Torres, A. (2019). Evaluation of a Wearable Device to Determine Cardiorespiratory Parameters From Surface Diaphragm Electromyography. *IEEE Journal of Biomedical and Health Informatics*, *23*(5), 1964–1971. https://doi.org/10.1109/JBHI.2018.2885138

Rios, E. A., Lai, C.-C., Yan, B.-R., & Lai, B.-C. (2021). Parametric Study of Performance of Remote Photopletysmography System. *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–4. https://doi.org/10.1109/ISCAS51556.2021.9401620

Rouast, P. V., Adam, M. T. P., Chiong, R., Cornforth, D., & Lux, E. (2018). Remote heart rate measurement using low-cost RGB face video: A technical literature review. *Frontiers of Computer Science*, *12*(5), 858–872. https://doi.org/10.1007/s11704-016-6243-6

Rudokaite, J., Ong, L. S., Janssen, M. P., Postma, E., & Huis In ’T Veld, E. (2022). Predicting vasovagal reactions to a virtual blood donation using facial image analysis. *Transfusion*, *62*(4), 838–847. https://doi.org/10.1111/trf.16832

Sabokrou, M., Pourreza, M., Li, X., Fathy, M., & Zhao, G. (2021). Deep-HR: Fast heart rate estimation from face video under realistic conditions. *Expert Systems with Applications*, *186*, 115596. https://doi.org/10.1016/j.eswa.2021.115596

Shimmer3 GSR+ Unit. (n.d.). Shimmer Wearable Sensor Technology. Retrieved May 11, 2023, from https://shimmersensing.com/product/shimmer3-gsr-unit/

Shimmer3 GSR+—IMotions. (n.d.). Retrieved March 2, 2023, from https://imotions.com/products/hardware/shimmer3-gsr/

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing*, *3*(1), 42–55. https://doi.org/10.1109/T-AFFC.2011.25

Stricker, R., Müller, S., & Gross, H.-M. (2014). Non-contact video-based pulse rate measurement on a mobile service robot. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 1056–1062. https://doi.org/10.1109/ROMAN.2014.6926392

Sun, W., Sun, Q., Sun, H.-M., Sun, Q., & Jia, R.-S. (2023). ViT-rPPG: A vision transformer-based network for remote heart rate estimation. *Journal of Electronic Imaging*, *32*(2), 023024. https://doi.org/10.1117/1.JEI.32.2.023024

Trost, Z., Jones, A., Guck, A., Vervoort, T., Kowalsky, J. M., & France, C. R. (2017). Initial validation of a virtual blood draw exposure paradigm for fear of blood and needles. *Journal of Anxiety Disorders*, *51*, 65–71. https://doi.org/10.1016/j.janxdis.2017.03.002

Tsouri, G. R., & Li, Z. (2015). On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras. *Journal of Biomedical Optics*, *20*(4), 048002. https://doi.org/10.1117/1.JBO.20.4.048002

van der Kooij, K. M., & Naber, M. (2019). An open-source remote heart rate imaging method with practical apparatus and algorithms. *Behavior Research Methods*, *51*(5), 2106–2119. https://doi.org/10.3758/s13428-019-01256-8

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

van Gent, P., Farah, H., Nes, N., & van Arem, B. (2018). Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data. *Proceedings Of*. https://repository.tudelft.nl/islandora/object/uuid%3A5c638e14-d249-4116-aa05-2e566cf3df02

Van Rossum, G. (2020). The Python Library Reference, release 3.8.2. *Python Software Foundation.*

Verkruysse, W., Svaasand, L. O., & Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. *Optics Express*, *16*(26), 21434. https://doi.org/10.1364/OE.16.021434

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, W., den Brinker, A. C., Stuijk, S., & de Haan, G. (2017). Algorithmic Principles of Remote PPG. *IEEE Transactions on Biomedical Engineering*, *64*(7), 1479–1491. https://doi.org/10.1109/TBME.2016.2609282

Wang, W., Stuijk, S., & de Haan, G. (2016). A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation. *IEEE Transactions on Biomedical Engineering*, *63*(9), 1974–1984. https://doi.org/10.1109/TBME.2015.2508602

What is RPPG? | Noldus. (n.d.). What Is RPPG? | Noldus. Retrieved May 2, 2023, from https://www.noldus.com/blog/what-is-rppg

Yu, Z., Li, X., & Zhao, G. (2021). Facial-Video-Based Physiological Signal Measurement: Recent advances and affective applications. *IEEE Signal Processing Magazine*, *38*(6), 50–58. https://doi.org/10.1109/MSP.2021.3106285

Zulko. (2017). *User Guide—MoviePy 1.0.2 documentation*. https://zulko.github.io/moviepy/index.html

Appendix A

Table 3: Overview of rPPG models since 2018

Name	Jaar	Short description	Performance	Public code	Ref	Include?
End-to-end systems						
DeepPhys	2018	DeepPhys consist of a forward convolutional neural network (2D-CNN). To process motion more efficiently, an algorithm based on skin reflection is represented.	Performed better than older methods in the (Chen & McDuff, 2018) paper, but does not perform better as tested in (Yu et al., 2023)	yes	(Chen & McDuff, 2018)	No
PhysNet	2019	PhysNet is an end-to-end model, which takes the RGB values of the face and maps these into a rPPG output.	Compared to DeepPhys, STVEN-rPPGNet, IPPG-3D-CNN, Physnet performs the best.(Ni et al., 2021) Also performs well on the PURE dataset (Sun & Li, 2022), but does not perform better as tested in (Yu et al., 2023)	yes	(Yu et al., 2019)	No
STVEN+rPPGNet	2019		Does not give better results than other models evaluated in (Sun & Li, 2022)		(Cheng et al., 2021)	No
AutoHR	2020		does not perform better as tested in (Yu et al., 2023)		(Yu et al., 2020)	No

Meta-rPPG	2020	Employs a meta-learner	Lower SD, MAE and RMSE than PhysNet, DeepPhys (and older methods). Performs best on the MAHNOB dataset as tested in (Yu et al., 2023)	yes	(Lee et al., 2020)	No
PyVHR	2020	PyVHR is a Python based framework, which uses nine different rPPG models within its pipeline. The nine traditional models are: ICA, PCA, GREEN, CHROM, POS, SSR, LGI, PBV, OMIT and a deep learning based method called MTTS-CAN.	PCA, SSR, POS and CHROM are the best models according to (Boccignone et al., 2020) However most articles do not explicitly describe the performance.	yes	(Boccignone et al., 2022)	Yes
Pulsenet	2022	This method is based on spatiotemporal convolution. It presents a robust heart rate by limiting the average bpm and PPGI. To suppress noise in the data, skin segmentation and an attention mechanism was introduced.	Has a pearson coefficient of 0.8. MAE 6.51. (on dataset PURE). Therefore better than deephys		(Yin et al., 2022)	No
VideoTransformer	2022		Does not perform better as tested in (Yu et al., 2023)		(Revanur et al., 2023)	No
PhysFormer	2023		Performs better than most other of the 14 tested methods on several datasets	yes	(Yu et al., 2023)	Yes

Other (not end to end) methods

		Model				
RythmNet	2019	Spatiotemporal map based, CNN-RNN model (trained to incl diverse illumination and pose)	Does not perform better on VIPL-HR, but has good correlation score on MAHNOB dataset, as tested in (Yu et al., 2023)	yes	(Niu, Shan, et al., 2020)	No
ST-Attention	2019		Does not perform better		(Niu et al., 2019)	No

			as tested in (Yu et al., 2023)			
PP-net	2020	Deep learning model, Long-term recurrent convolutional network.	Perform best on most datasets as tested in (Sun & Li, 2022; Yu et al., 2023)	Yes	(Panwar et al., 2020)	Yes
NAS-HR	2021		Does not perform better as tested in (Yu et al., 2023)		(Lu & Han, 2021)	No
CVD					(Niu, yu, et al., 2020)	No
Contrast- Phys	2022	Unsupervised, 3DCNN, using contrastive loss	Has good scores on correlation on most datasets	yes	(Sun & Li, 2022)	No
(nameless)	2022	CNN + LSTM	Good recording the authors.	?	(Yen et al., 2022)	No

Appendix B

Figure 5

Overall extracted heart rate using LGI, POS, and PCA models compared with the Shimmer heart rate for participant 6.

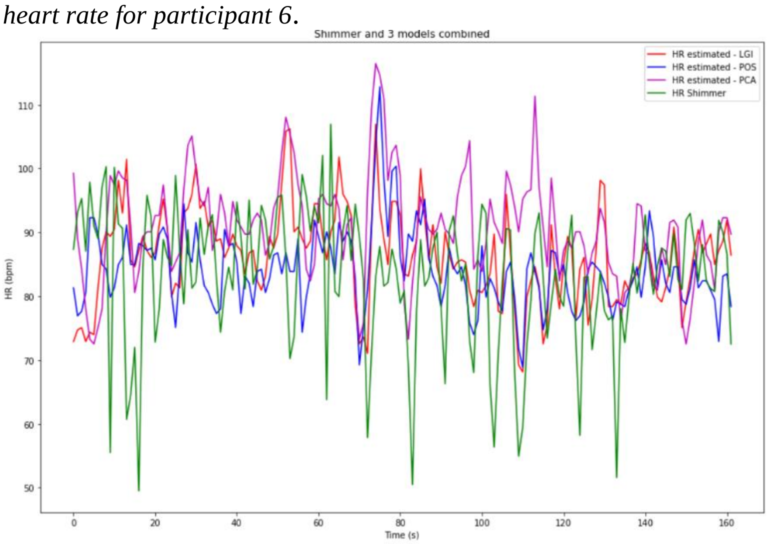


Figure 6

Overall extracted heart rate using LGI, POS, and PCA models compared with the Shimmer heart rate for participant 11.

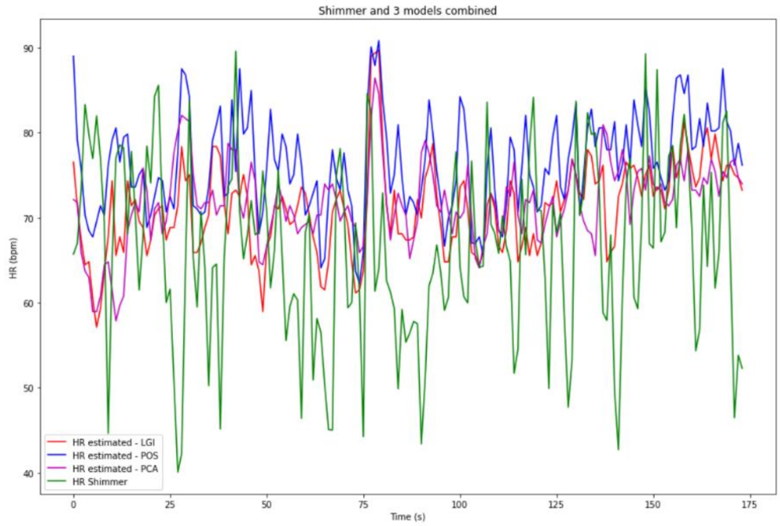


Figure 7

Overall extracted heart rate using LGI, POS, and PCA models compared with the Shimmer heart rate for participant 39.

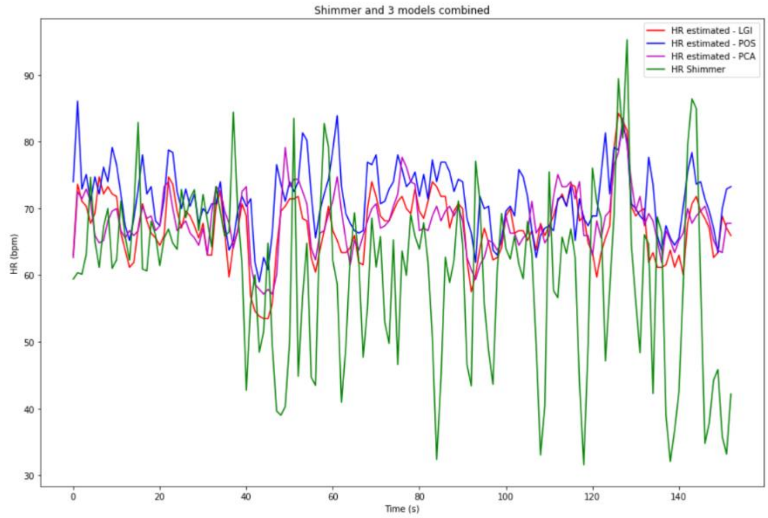


Figure 8

Overall extracted heart rate using LGI, POS, and PCA models compared with the Shimmer heart rate for participant 44.

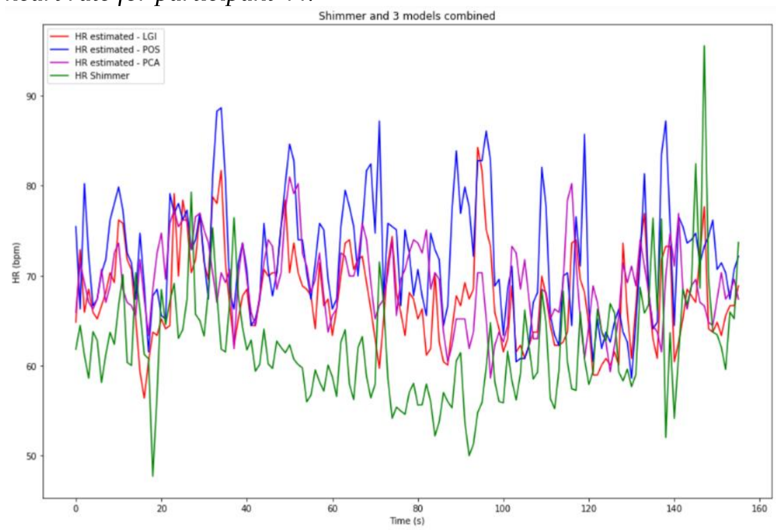


Figure 9

Overall extracted heart rate using LGI, POS, and PCA models compared with the Shimmer heart rate for participant 52.