11-2019

# VitaMon: Measuring heart rate variability using smartphone front camera

Sinh HUYNH
*Singapore Management University*, npshuynh.2014@phdis.smu.edu.sg

Rajesh Krishna BALAN
*Singapore Management University*, rajesh@smu.edu.sg

JeongGil KO
*Yonsei University*

Youngki LEE
*Singapore Management University*, YOUNGKILEE@smu.edu.sg

# VitaMon: Measuring Heart Rate Variability Using Smartphone Front Camera

Sinh Huynh
Singapore Management University
npshuynh.2014@smu.edu.sg

Rajesh Krishna Balan
Singapore Management University
rajesh@smu.edu.sg

JeongGil Ko
Yonsei University
jeonggil.ko@yonsei.ac.kr

Youngki Lee
Seoul National University
youngkilee@snu.ac.kr

## ABSTRACT

We present *VitaMon*, a mobile sensing system that can measure the inter-heartbeat interval (IBI) from the facial video captured by a commodity smartphone's front camera. The continuous IBI measurement is used to compute heart rate variability (HRV), one of the most important markers of the autonomic nervous system (ANS) regulation. The underlying idea of VitaMon is that video recording of human face contains *multiple* cardiovascular pulse signals with *different* phase shift. Our measurement on 10 participants shows the significant time delay (36.79 ms) between the pulse signals measured at the jaw region and forehead region. VitaMon leverages deep neural network models to extract both spatial and temporal information of the video to reconstruct a pulse waveform signal that is optimized for estimating IBI. We evaluated VitaMon with a dataset collected from 30 participants under various conditions involving different light intensity levels and motion artifacts. With the 15 fps video input (66.67 ms time resolution), VitaMon can measure IBI with an average error of 14.26 ms and 21.65 ms using personal and general model respectively. HRV features including geometry Poincare plot, time- and frequency-domain features extracted from the IBI measurement all have high correlation with the reference signal.

## CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Empirical studies in ubiquitous and mobile computing**.

## KEYWORDS

Heart Rate Variability, Photoplethysmography (PPG), Remote PPG, Mobile Sensing

**ACM Reference Format:**
Sinh Huynh, Rajesh Krishna Balan, JeongGil Ko, and Youngki Lee. 2019. VitaMon: Measuring Heart Rate Variability Using Smartphone Front Camera.


*SenSys '19, November 10–13, 2019, New York, NY, USA*

ACM ISBN 978-1-4503-6950-3/19/11…$15.00
https://doi.org/10.1145/3356250.3360036

In *The 17th ACM Conference on Embedded Networked Sensor Systems (SenSys '19), November 10–13, 2019, New York, NY, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3356250.3360036

## 1 INTRODUCTION

Heart rate variability (HRV), fluctuations in the interval between consecutive heartbeats, is an important physiological marker that reflects the changes in the sympathetic-parasympathetic balance of the autonomic nervous system (ANS). HRV has proven its effectiveness as a diagnostic tool in various research and clinical studies related to cardiovascular disease, diabetic autonomic dysfunction, hypertension, and psychiatric and psychological disorders. Furthermore, daily HRV monitoring could be useful for screening and tracking the condition of individuals at risk to serious health issues [15, 44, 49, 52, 61]. More generally, beyond just clinical use, HRV measurements can also help to measure stress and engagement levels of a person performing various tasks, and it can also help monitor sleep quality [53, 55].

Conventional HRV measurement techniques, however, have two significant drawbacks to be used for daily HRV monitoring. First, they require additional electronic or optical sensing devices [8, 24, 50] that are often not available to most people. Second, most sensing instruments need direct contact to the skin for reliable signal acquisition, making daily continuous measurements tedious and uncomfortable. For example, electrocardiograph (ECG) recording devices require several electrodes to be carefully attached to different body points making it impractical as a general daily use solution. Recently, more practical photoplethysmogram (PPG)-based techniques are available that measure cardiac activities based on video recordings of a finger or a human face. However, most of them are limited to detecting just the heart rate (HR) [32, 39, 46, 62].

In this paper, we propose a novel contactless HRV sensing system, named *VitaMon*, to measure HRV from a video of the user's face captured by a smartphone front camera. Our system has two clear benefits over prior techniques: (1) *VitaMon* does not require any extra sensing device and uses a commodity smartphone's front camera, possibly with low resolution and frame rate, (2) sensing can be done naturally and unobtrusively while the user uses the phone for different purposes (e.g., plays games or video-chat). This opens up new opportunities to apply real-time HRV sensing in mobile apps – e.g. to track the stress and engagement levels of users playing a mobile game or using an education app.

Building *VitaMon* required solving several challenges before measuring HRV using front-camera videos became viable. The core of

HRV monitoring is to calculate the precise intervals between the two peaks of consecutive heartbeat cycles. Techniques have been proposed to count the number of heartbeat peaks (HR) from the changes of the reflected light intensity in the video recordings, but it has still not been feasible to identify the exact peak times. The reflected light signals captured in video recording often have unclear peaks due to noise, different ambient light conditions, and motion artifacts (head, face and hand movement), making it difficult to accurately detect the peaks. Also, the low frame-rate of the front camera (e.g., 15 fps) makes it difficult to estimate the exact times of heartbeat peaks as the peaks may occur in between two consecutive video frames.

*VitaMon* addresses these challenges using two key insights. Firstly, *VitaMon* takes multiple PPG readings from different facial areas in a single frame whereas prior camera-based PPG techniques consider facial video as a single image. Multiple facial regions carry pulse signals from the heart at different time offsets and shifted phases, which enables *VitaMon* to overcome limitations from low frame rates and noise. Second, we observed that there is a strong temporal correlation between PPG signal patterns and ECG signal patterns. *VitaMon* utilizes the correlation to build a deep learning model that generates exact heartbeat peaks from low-quality PPG signals.

We built a novel HRV estimation technique based on the above two insights. The technique is designed with a two-stage Convolutional Neural Network (CNN). The first network learns the correlation between the ECG signals and the PPG signals (estimated from the video), and firstly reconstructs a form of ECG waveform from the captured video to identify which video frame includes a peak. The second CNN learns the relationships between the facial images (the reflected light intensity of the multiple facial regions) and the temporal distance between the actual peak time and the image capture time. Based on the trained model, *VitaMon* estimates the exact timestamp of the peak.

The contributions of this work are as follows:

- We design *VitaMon* a contactless HRV monitoring system using videos of user's face captured by a commodity smartphone's front camera with low frame rate.
- Our motivational study shows that PPG-based heartbeat estimation with facial videos can achieve higher granularities than the video's frame rate. Such fine grain measurements allow the detection of heartbeat intervals at millisecond-level accuracy.
- We built a novel HRV estimation technique based on Convolutional Neural Networks (CNN) that can accurately estimate the exact timestamps of heartbeat peaks from a facial video.
- We evaluate *VitaMon* with data collected from 30 participants under different smartphone usage conditions. The results show that our technique can detect heartbeat intervals only with 14.26 ms of errors. Also, it is robust against the light conditions and motion artifacts. Finally, through a user study, we show that *VitaMon* can be used in various practical applications such as stress detection.

## 2 BACKGROUND AND RELATED WORK

Cardiac activity monitoring is the basis of many clinical, healthcare and psychological condition monitoring applications. In particular, heart rate variability (HRV) measurements can be used for many applications from early-warning of impending cardiac disorders, diagnosing various diseases, to activity-associated stress monitoring. While HRV can be captured in several different ways, in this work, we focus on the use of the photoplethysmogram (PPG) measurement, a low-cost and easy-to-apply method for measuring heartbeat. Nevertheless, since most clinical-grade PPG sensors still require a physical attachment of the special sensor, there has been prior work to alleviate such inconvenience and use camera images for PPG monitoring [19, 31, 38, 45, 62]. In this section, we present background information on how PPG sensors work and discuss how previous work utilizes camera-captured data to design non-invasive systems for measuring HRV.

### 2.1 Photoplethysmogram (PPG)

Photoplethysmogram (PPG), initially developed in the 1930s, is an optical sensing technique for detecting heart pulse [20]. It is based on the principle that blood absorbs light more than the surrounding tissue; thus, variations in the blood volume will affect the transmission or reflectance of light correspondingly. The conventional design of a PPG sensor includes a light emitting diode (LED) to illuminate a region on the skin and a photodiode to measure the intensity of the reflected light. This light intensity is inversely related to the blood volume, therefore, the pulsatile component of the PPG signal oscillates with every heartbeat cycle.

Being an easy-to-use, low cost, and convenient sensing technique for understanding cardiac activities, PPG sensing technology has been extensively studied, and have continuously improved over time to the point where the accuracy of these measurements can be used to compute HRV. However, despite its accuracy, the fact that PPG measurements typically need a sensor continuously attached to the skin limits the realization of ubiquitous monitoring applications [58].

### 2.2 Other HRV Monitoring Techniques

Aside from PPG-based HRV monitoring, sensors of different modality such as electrical, acoustic, seismic sensors can be used to measure the inter-beat interval of heartbeat, and this information can be used to interpret HRV [12, 13, 65]. However, given that they require cumbersome attachments to the skin for accurate measurements, the HRV captured from these devices are mostly used within clinical environments [4]. Nevertheless, with recently introduced wearable and mobile ECG monitors, there has been a number of efforts in measuring HRV from mobile devices. The work by Nepi et al. compared the performance of a Zephyr Bioharness ECG sensor to clinical-grade devices to validate their clinical effectiveness [41]. Wippert et al. show performance evaluations of different mobile ECG platforms for detecting various cardiac activity features, which include HRV [64]. While these work along with may similar efforts [17, 27, 36] show promising results for its applicability in various domains, usability issues with these devices yet remain [63].

### 2.3 Remote PPG

*Remote PPG* techniques, which do not involve a sensor attached to the skin, have been recently proposed to improve the convenience of PPG measurement for daily monitoring. This body of work utilizes a camera to capture subtle changes in the skin color as the pulse wave

| Related Work | N | Video Input | Additional Processing Steps | Ref. Signal | Result (MAE) |
|---|---|---|---|---|---|
| [5] | 5 | 30fps, 800x600 | Fuse with additional ballistocardiographic signal | ECG | 24.4 ms |
| [14] | 20 | 60fps, 720x480 | Extract the interval from the pulse signal's first derivative | ECG | 35.3 ms |
| [39] | 14 | 30fps, 960x720 | Filter out pixels covering eye area<br>Apply ICA on color channel signals to recover source signals | Contact PPG | 26 ms |
| [47] | 15 | 15/60fps, 1280x720 | Filter out pixels covering hair, beard, eyebrows area | Contact PPG | 15.04ms |

**Table 1: Related work on non-contact HRV measurement in stationary condition using signal processing approach. N: sample size of the study; MAE: Mean Absolute Error.**

propagates from the heart through the body. This color change is not visible with the human eyes but can be captured using an RGB camera. For instance, Poh et al.[45] is one of the early works that introduced a remote heart rate assessment technique using a webcam under ambient light conditions, showing the potential to apply these techniques in various applications. The main processing pipeline in the proposed techniques generally includes the following steps: (1) detecting face region in each input frame; (2) averaging the pixel value of the face region in consecutive frames to reconstruct the pulse signal; (3) Up-sampling or applying interpolation and bandpass filter on the pulse signal; (4) Performing peak detection and count the number of heartbeat. More recent works exploit the smartphone's camera to implement remote PPG using similar processing approach. In particular, Kwon et al.[32] has demonstrated the feasibility of using smartphone camera (iPhone) to estimate heart rate from facial video recording and reported the error of 1.08% (beat per minute).

While being attractive for heart rate monitoring, remote PPG schemes proposed until now are inefficient for HRV monitoring. A primary reason is that HRV monitoring requires 100 Hz sampling rate [56], whereas smartphone cameras operate at a slower sampling rate (e.g., 15 Hz for smartphone front cameras). If the camera captures videos at 15 fps, the granularity or time resolution of the heartbeat peak detection would be 67 *msec* in the ideal case. Furthermore, external lighting conditions and motion artifacts (especially in mobile context) would further complicate the process of capturing accurate PPG measurements. While some work suggests the use of signal processing to overcome these challenges [31, 33], the low frame rate and resolution of the smartphone's front camera still heavily impact the sensitivity of HRV measurements [56]. For instance, as shown in Table1, Davila et al.[14] showed that with a similar processing pipeline mentioned above, they can perform the IBI measurement with 35.3ms Mean Absolute Error (MAE) . While Rodriguez et al. [47] reported a more promising result IBI estimation (MEA 15.04ms) using a similar processing technique, it is worth noting that the IBI estimation was evaluated against another finger pulse sensor. Previous studies have shown that the interval measured by pulse sensor even with sampling rate as high as 1kHz may still have certain error compared to the interval extracted from ECG signal [14, 24]. This is due to the difference of the two signal waveforms, the peak of PPG waveform is not as distinctive as the R-peak in ECG waveform.

Different from these previous works, we empirically show that the time delay of pulse signal traveling through facial regions is

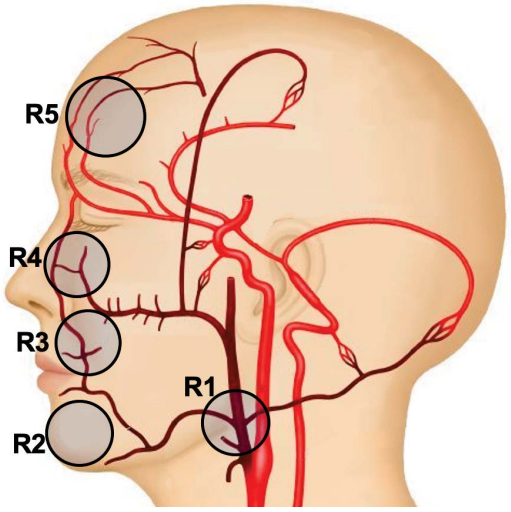



**Figure 1: Anatomy of facial artery (This figure is drawn based on [1]).**

significant as compared to the time resolution of commodity camera. This finding suggests that a video of human face is a source of multiple signals of blood volume pulse with different phases or time delays. We propose an approach using convolution neural network to leverage such spatial-temporal information from the input video to estimate the IBI with higher precision.

## 3 INVESTIGATION: CAN YOU EXTRACT MULTIPLE PPG DATA POINTS FROM FACIAL IMAGES?

The basis of this work is on an important hypothesis: "Given the structure of facial arteries, different parts of the face will show PPG 'peaks' at different times." We can exploit this information to gain more precise peak-occurrence times, finer than the frame rate granularity.

Prior works introduce the concept of *pulse transit time*, the time a pulse wave to travel between two arterial sites [16, 54]. Existing measurements from the heart to ear and finger [23] has shown that the pulse transit times from the heart to the ear and finger is ∼174 ms and ∼245 ms, respectively. We hypothesize that it would take some

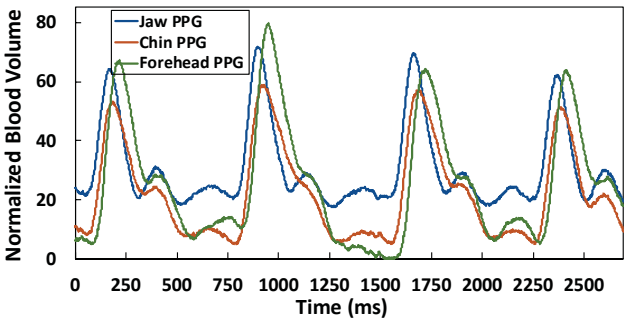



**Figure 2: Normalized PPG signals at different facial positions.**

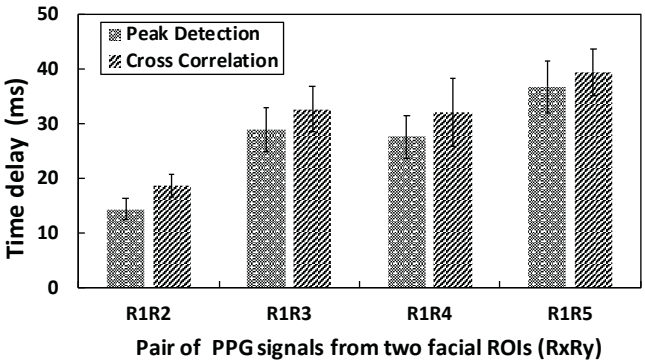



**Figure 3: Time delay of PPG peaks between two different facial positions.**

time (significant compared to the time resolution of a video) for the pulse to propagate even with the facial arteries. To the best of our knowledge, no previous work quantifies the time delay of the pulse traveling on different facial regions. Instead, previous work either neglect this time delay or make assumptions that within the face, the time delay is not significant.

To validate our hypothesis, we conducted an IRB-approved preliminary study with 10 participants (ages from 19 to 31, 4 females). We selected five facial regions and attach a photoplethysmogram (PPG) sensor to each of these regions to understand how the PPG-peak delay occurs for different regions. The five facial regions, $R1$ jaw corner, $R2$ center chin, $R3$ upper lip, $R4$ below left eye, and $R5$ forehead, were selected based on the anatomy of facial arteries as illustrated in Figure 1. Participants in this study were asked to sit on a chair while the PPG sensor captures samples at 1 $kHz$ for one minute. All five PPG sensors were attached to an Arduino, and the five incoming signals were time-synchronised.

Figure 2 shows an example of normalized PPG signals from our collected data. We can observe a phase shift of the peak of the PPG signals detected at different locations. This observation suggests that we can exploit this spatial-temporal aspect of PPG signals, based on the artery structures of a person's face. We further quantify the time difference for signals observed at two different facial regions in two ways: (a) using peak detection and (b) phase-shift calculation via cross-correlation computation [6, 30]. Figure 3 shows the results. The time delay for different facial region pairs are as significant as $\sim$36.79$msec$, when the pulse travels from the corner jaw to the forehead. The delays are consistent over the two quantification methods.

Potentially, as we will detail in the following section, *VitaMon* exploits this to make very accurate measurements of the heartbeat interval even with videos taken at low frame rates.

## 4 DESIGN OF *VITAMON*

*VitaMon* measures a user's heart rate (HR) and heart rate variability (HRV) using just videos captured from that user's front facing phone camera by exploiting the color changes that occur as blood pass through the facial arteries. Figure 4 shows *VitaMon*'s data pipeline and we explain each stage in more detail next:

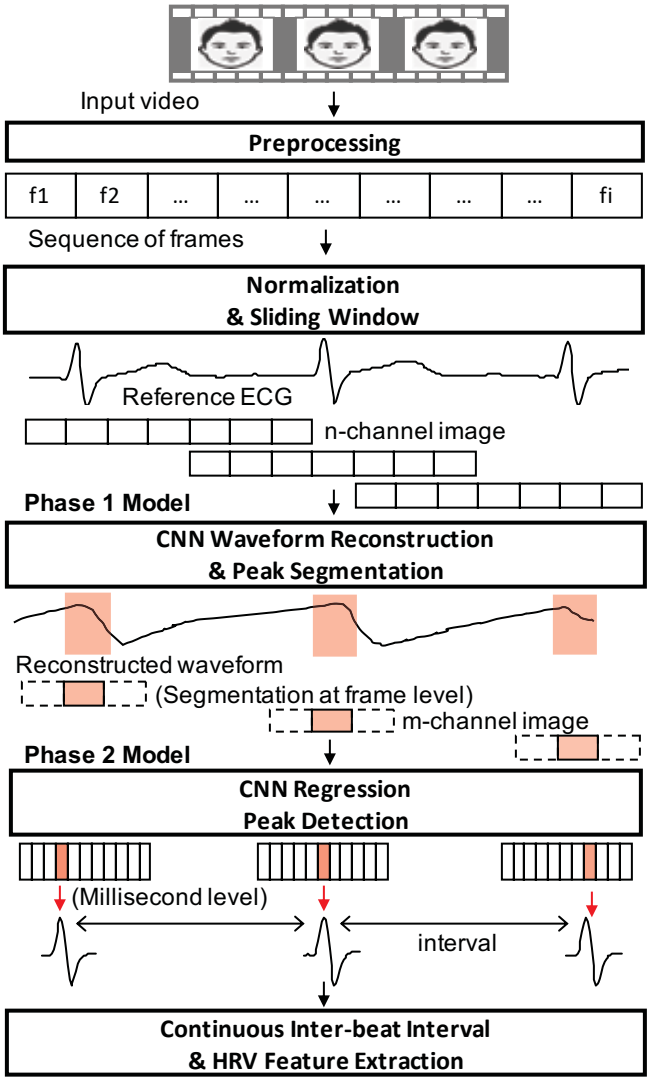



**Figure 4: VitaMon data pipeline.**

### 4.1 Preprocessing: Extract the Green Color Channel

Starting from the topmost preprocessing phase, we first resize each frame from the videos to 224x224 resolution, then extract and normalize the green color channel of each frame from the videos captured by a smartphone's front camera. The key principle used by *VitaMon* is that blood absorbs light more than the surrounding tissues in the body and that the absorption levels are directly proportional to the blood volume [62]. This phenomenon causes subtle color changes to appear on human skin, which are invisible to human eyes but can be captured by camera images. Prior work has shown that the green channel captured by RGB camera is better than red and blue channels in detecting these colour changes [31, 62]. This is because the absorption spectra of hemoglobin (Hb) and oxyhemoglobin (HbO2), the two main constituent components of blood, peaks in the 520 to 580 nm light spectrum – which falls in the middle of the green spectrum [62]. Thus any changes in the blood volume, caused by heartbeats etc. will be easier to detect using the green channel information compared to the other colours.

### 4.2 Normalisation & Input Creation

*VitaMon* processes the green channel information to predict the HR and HRV of the person in the captured images. However, processing every frame produced by the camera is computationally very expensive. Thus, *VitaMon* creates a multi-channel image that is formed by stacking multiple green color-channels extracted from consecutive video frames which is used as the input for subsequent machine learning stages. In particular, we extract green channel samples in sets of $n$ samples to form a single image that combines the features contained in the $n$ samples. By doing this, the depth dimension of this stacked image will contain the temporal information of $n$ consecutive green frames. We found, empirically, that $n = 25$ worked best for 15 fps video feeds – with each stack containing 25 samples representing changes in the green channel over a period of 1.67 seconds. This is sufficiently long to allow us to detect a full heartbeat cycle, even for heart rates as low as 36 bpm, just from a single image.

This stacking serves three main purposes: (1) it reduces the input size to minimize model complexity; (2) stacking a single color channel to form an image allows the depth dimension of the image to contain the temporal aspects – this separates away the color/spectral information making the technique much more robust; (3) we now have a single image that contains *both* spatial and temporal information of the facial video, allowing us to extract pulse information from the image using just a single 2D convolution.

### 4.3 Two-Phase Machine Learning

Reliable HRV measurement requires accurate identification of the R-peaks of the ECG and their occurring timestamps in the cardiovascular pulse signals generated as the heart pumps blood around the body. This is different from just measuring the heart rate as heart rate calculation uses an average of the number of beats over a minute (bpm) while HRV measures the inter-beat time in milliseconds.

To effectively extract the HRV using just video images of a user's face, *VitaMon* uses two phases: (1) it reconstructs the "frame-order waveform" of the ECG signal to identify heartbeat cycle peaks from the video sequences, and (2) it then estimates the exact timestamp of each peak. In both phases, ECG reference signal is used only for the purpose of model training and evaluation; *VitaMon* only takes the facial video as input.

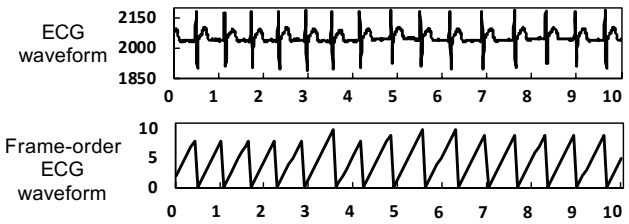



**Figure 5: Example of ECG and frame-order ECG waveforms.**

### 4.4 Phase 1: Reconstruction & Segmentation

Prior work has used photoplethysmogram (PPG)-based methods to reconstruct the blood volume signal directly from video recordings. However, these methods are limited to just detecting the HR and achieved poor results when used to also detect the HRV. To detect the HRV, we use a CNN-based regression model with the Inception module from InceptionV3 model [57]. We add enhancements to reconstruct a pulse waveform in frame-units by utilizing the color changes embedded in the 25-channel stacked input images described earlier. Specifically, as each stacked image holds the facial color change information (on the green channel) for at least one full heartbeat cycle, we trained the CNN model to identify the exact sub-frames within this stacked image where the heartbeat cycle's "peaks" have taken place. We did this by using the intuition that the this peak will cause a noticeable color change (on the camera) due to large amounts of blood flowing through the arteries.

Based on this, we mark the center-most frame in the $n$-channel image with the respect to the nearest peak that occurs earlier. For instance, if the peak occurs at the $13^{th}$ sub-frame (e.g., the center-most frame of the 25-channel image), the model will output a value '0'. If the peak occurs on sub-frame 10, three sub-frames before the center, the model will output '3'. We label the data by marking the offset of the center frame according to only the previously observed peak. Hence, the offset of the center frame is always a positive number. Using this simple scheme, for each of the 25-channel images, we can identify at what location (in the units of sub-frames/channels) the heartbeat's peak occurred with reference to the most recent peak. When done for all of the image sequences, we can construct a "frame-order waveform", which is roughly correlated to the ECG at a frame-level granularity. Figure 5 illustrates an example of the ECG waveform and its corresponding frame-order waveform.

There are two major benefits of using this approach. First, the frame-order waveform represents a normalized form of ECG. Naturally, using the frame-order waveform eliminates the effects caused from ECG peak amplitude variations and facilitate the model to focus on the local relative change of the blood volume within the samples. Using this information, our CNN model is optimized to learn while focusing on the differences in color distributions among different neighboring frames. Second, the use of the frame-order waveform allows the model to easily distinguish between two consecutive heartbeats (i.e., the end of one heartbeat cycle and the beginning of the next). Once the value decreases, we can quickly

notice that one heartbeat cycle has ended. When using a PPG-based approach, due to the smooth signal patterns, making this distinction of whether the currently detected sample is before or after a peak is difficult. By applying the frame-order waveform, the model, in its training phase, can penalize heavier if an estimation is made for a different subsequent heartbeat cycle.

Note that the range of the frame-order waveform will vary with respect to the inter-heartbeat-interval, given that this interval varies from person to person in the typical range of 500 ms to 1470 ms under resting condition [48]. For instance, for a person with the a heartbeat interval of 600ms, the output values will range from 0 to 9 (in unit of frames; assuming 15 fps), while a person with 1000 ms interval, the output values will vary from 0 to 15.

### 4.5 Phase 2: Peak Detection

Next, in the second phase of our model, we take the $n$-channel images that are labeled from the first CNN model as '0' (i.e., the peak has occurred at the center-most sub-channel for the image) and cut-off sub-channels on the edges in a symmetric manner. As a result, we leave only the $m$-channels in the center of the original $n$-channel image (where $m < n$), and maintain the peak-detected channel at the center of the stack. We then train a second CNN-based regression model using these images and the ground truth ECG waveform. By doing so, we can now correlate the ECG peaks with the exact location of where *within* the peak-detected channel the R-peak took place. Given that the color distribution for the peak-detected channel will vary for different (more specific) R-peak occurrence locations, we can start making fine-grain estimations (at the msec-level) on the actual time that the R-peak occurred at a granularity finer that that of the frame rate. Again, this is based on the findings from our preliminary studies indicating that the pulse will travel at slow speeds even within a person's facial regions. Meaning that for some images, we will have the peak at the jaw region of the face, and for some, the peak will be at the forehead. Each of these images will have different points at which the ground-truth ECG presents its R-peak. Learning this information is the core of this second phase CNN design.

### 4.6 VitaMon Implementation

We implemented *VitaMon* as an Android application with the phase 1 and phase 2 models implemented in *tflite* format with float32 precision. Figure 6 presents the overall CNN structure of both our phase 1 and phase 2 models. Note: each convolution layer is followed by a batch normalization layer and a rectified linear unit (ReLU) activation layer. In terms of model complexity, the total number of parameters is 508,129 and 104,129 for the Phase 1 and Phase 2 models, respectively. The complexity did not increase by using stacked images as a similar schema for processing a standard 3-channel RGB image would have 503,233 and 102,689, parameters for the two phases, respectively.

We ran *VitaMon* on different octa-core phone devices including the Lenovo Phab 2 (2016), Galaxy S8 (2017), and Huawei P20 Pro (2018). Table 2 reports the running time of *VitaMon*'s main processing components on each phone using just CPU resources (no GPU optimisations done yet). The preprocessing step in Table 2 refers to the process of extracting the green channel from each video frame and creating the stacked images that are fed as inputs to the CNN models.

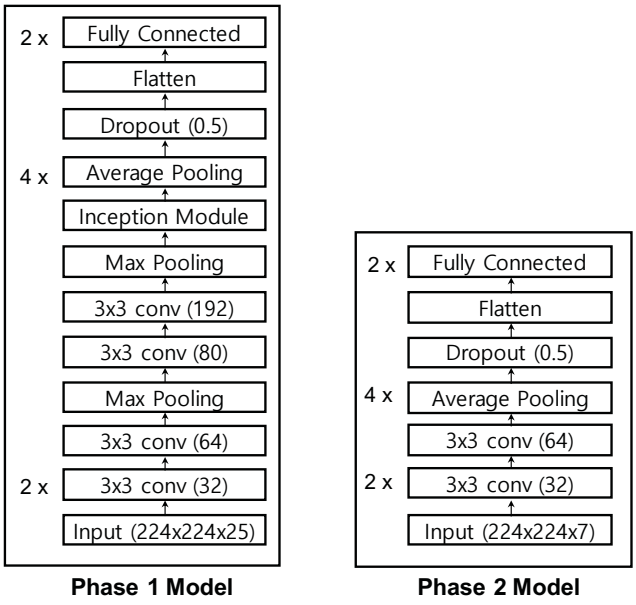



**Figure 6: Structures of phase 1 and phase 2 models.**

| | Processing time (ms) | | |
|---|---|---|---|
| Device | Pre-process | Phase1 model | Phase2 model |
| Lenovo Phab 2 | 31.4 | 122.2 | 45.4 |
| Galaxy S8 | 5.6 | 125.6 | 47.1 |
| Huawei P20 Pro | 8.5 | 108.4 | 44.4 |

**Table 2: Operation latency of *VitaMon*'s components on three different mobile devices.**

The operational latency of the phase 1 model is longer compared to the phase 2 model as it is a more complicated model (as described earlier). Overall the latency of *VitaMon* is sufficient for real-time use. We can improve the latency further, as future work, by further optimising the models or by using pruning or quantization scheme [18] and/or GPU optimised DNN runtimes such as DeepMon [22] or TensorFlow Lite running on mobile GPU [59].

## 5 DATA ACQUISITION

### 5.1 Sensors and Set-up

In this study, we use a Lenovo Phab Pro2 smartphone to record the facial video of participants and a Zephyr Bioharness 3 ECG strap to acquire ground truth reference pulse signals (e.g., ECG). All videos were recorded using a frame rate of 15 fps with a pixel resolution of 1920x1080 from the 8-megapixel front camera with 3.75 mm focal length and a f/2.2 lens aperture. The automated white balance (AWB) mode of the camera was enabled to normalise the color representations of the captured images under different lighting

conditions. The ECG signal was recorded simultaneously throughout the experiment using the Zephyr ECG strap with a sampling rate of 250 Hz. The Zephyr is FDA-approved and multiple prior studies have used it to provide reference ECG/RR interval data under various conditions [25, 26, 28, 40].

## 5.2 In-lab Data Collection

We conducted an IRB-approved study with 30 participants of different ages (24 to 39) and skin tones (22 participants with light yellow skin tone from South East Asia and East Asia, 6 participants with dark brown skin tone from South Asia, and 2 participants with fairer skin tone from Europe). Participants were seated on a height-adjustable chair at a table in front of a tripod holding the smartphone mounted vertically. The distance from the smartphone's front camera to the participant's face varied from 25 to 50 cm depending on the participant's preferred sitting posture.

Each participant did eight 5-minute tasks that were fully recorded by the smartphone. Each of the eight tasks was designed to capture different motion artifacts and light conditions. The eight tasks were: tasks one to five required the participant to stay as still as possible the entire five minute duration with each task using a different fluorescent light intensity. The intensities used were 150 (denoted as $L1$), 250 ($L2$), 380 ($L3$), 600 ($L4$), and 1000 ($L5$) lux and represented different types of real-world intensities. For example, the recommended light level at homes is 150 lux, 500 lux for the library and 750 lux at supermarkets [42]. The last three task required the participant to perform an action under a consistent 380 lux lighting condition. The three tasks were ($M1$) Speaking: Counting out loud from 1 to 100 repeatedly. ($M2$) Horizontal head rotation: Participants had to rotate their heads horizontally by 120 degrees at a speed of about 20 degrees/sec. ($M3$) Manual phone holding: The smartphone was removed from the tripod and held by the participant in their hands, with the front facing camera still being able to see their faces, for 5 minutes.

## 5.3 Real-world Experiments

In addition to the controlled lab studies, to evaluate the robustness of *VitaMon* in real-life scenarios, we collected data from two participants while they performed various real-world tasks. In particular:

(1) Passenger in a driving car: This scenario introduces different types of motion artifacts as the car moves on the road (e.g., accelerate, slow down, stop, bumps at potholes). Figure 7 shows an example of the acceleration signals (excluding the gravity) collected from the phone that participants used to record the facial video. The light conditions also change dynamically as the car moves in and out of shaded and non-shaded areas. Each participant held the phone in their hand, with the front camera facing their faces, for two 5-minute sessions.

(2) Coffee shop: This scenario required each participant to record their faces for two 5-minute session while sitting in a very dim (40 lux) coffee shop.

# 6 EVALUATIONS

For the valuation of *VitaMon* , we train the model using two types of data and create two versions of *VitaMon* : (1) a *global* model with training data from multiple people, and (2) a *personalized* model trained with a specific person's previously collected data. For each case, we evaluate the accuracy of different metrics that can be extracted from a person's heart. Specifically, we focus on the accuracy of the heart rate, inter-beat-interval, and HRV. We evaluate the global model with the leave-one-out subject level evaluation and the personalized models with leave-one-out session level evaluation.

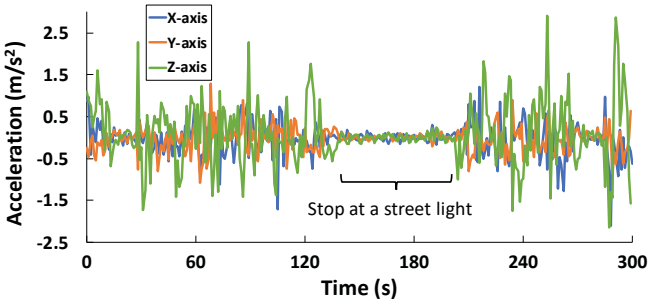



**Figure 7: Acceleration signals during the real-world experiment, passenger in a driving car scenario.**

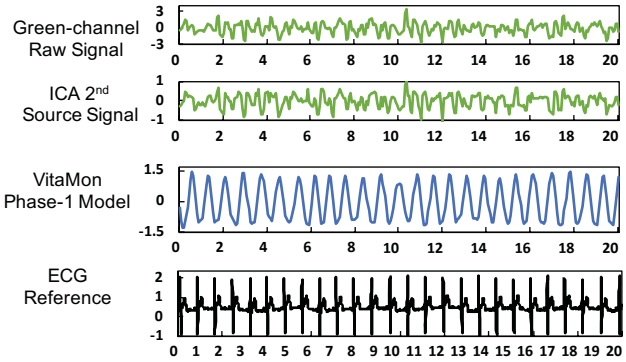



**Figure 8: Raw average signal extracted from green channel of whole face region; Second component of ICA; VitaMon phase-1 model output; ECG reference signal.**

## 6.1 Heart Rate Detection

We first evaluate the performance of *VitaMon* in calculating the heart rate from the captured video. *VitaMon* calculates the heart rate using the output of the Phase-1 model; it identifies a frame that includes the peak of a heartbeat and the heart rate can be calculated by simple counting of such peak frames. We used a 1-minute window to calculate the heart rate and slide the window every second. We also compare the results with the state-of-the-art signal processing-based remote PPG schemes as discussed in Section 2 [45, 46].

Figure 8 shows the waveforms reconstructed by *VitaMon*'s Phase-1 model for 20-second epoch signal from our dataset, along with the comparison with a state-of-the-art technique (a signal decomposition method based on Independent Component Analysis (ICA) [45, 46]) and the reference ECG signals. The third plot in the figure shows that our approach shows a clear representation of the pulsatile variations, closely correlated with the ground-truth ECG traces in the bottom-most plot. On the other hand, the ICA-based method results in a much unclearer waveform (as shown in the second plot), from

| | | Light Condition | | | | |
|---|---|---|---|---|---|---|
| Metric | Model | L1 | L2 | L3 | L4 | L5 |
| HR MAE | General | 0.82 | 1.06 | 0.82 | 0.94 | 0.88 |
| (bpm) | Personalized | 0.67 | 0.72 | 0.61 | 0.61 | 0.56 |
| Peak Position | General | 0.78 | 0.98 | 0.76 | 0.80 | 0.84 |
| MAE (frame) | Personalized | 0.63 | 0.72 | 0.65 | 0.72 | 0.62 |

**Table 3: Phase-1 model evaluation under different light conditions: Mean Absolute Error (MAE) for heart rate (HR) and peak position estimations. L1-L5 are set to 150, 250, 380, 600, and 1000 Lux, respectively.**

| | | Motion Artifact Condition | | | |
|---|---|---|---|---|---|
| Metric | Model | M0 | M1 | M2 | M3 |
| HR MAE | General | 0.82 | 1.77 | 1.69 | 1.31 |
| (bpm) | Personalized | 0.61 | 1.23 | 1.38 | 1.08 |
| Peak Position | General | 0.76 | 1.33 | 1.45 | 1.32 |
| MAE (frame) | Personalized | 0.65 | 1.02 | 1.19 | 1.18 |

**Table 4: Phase1 model evaluation under different motion artifact conditions: Mean Absolute Error (MAE) for heart rate (HR) and peak position estimations. M0-M3 are set to "no action", "speaking, "horizontal head rotation", "manual mobile phone holding", respectively.**

which heart rate calculation is not still trivial. The Phase-1 CNN model identifies peak frames from a noisy signal by leveraging the relationships between ECG signals and PPG signals whereas the signal reconstruction is not effective based on signal processing techniques.

We then quantitatively compute two metrics (1) the mean absolute error (MAE) for the estimated heart rate compared to the ground-truth, and (2) MAE for peak position (e.g., the peak position error is 1 if the 10th frame should have the peak but our Phase-1 model identifies 11th frame as the peak frame). Tables 3 and 4 show the results for different lighting conditions and motions artifacts, respectively. Each lighting condition $L1$-$L5$ and motion artifact $M0$-$M3$ correspond to the different conditions discussed in Section 5. We use $L1$ and $M0$ by default.

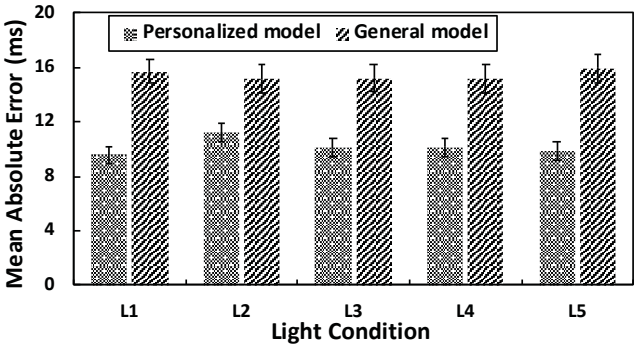
Table 3 presents that *VitaMon*, despite under different light intensities, have the exceptional performance of keeping HR estimation error under a single beat. The personalized model, as one may expect, outperforms the general model, but for both cases, the errors are kept extremely low. Overall, the CNN model we designed were robust against different light conditions, allowing a reliable heart rate measurement.

Results in Table 4 suggest that, with motion introduced, the heart rate estimations are affected more than simple light condition changes. Especially when parts of the facial components move (due to talking in $M1$) and the entire face rotates ($M2$) the error increases to higher than 1 bpm. Small variations due to hand-holding the smartphone ($M3$)) show relatively less loss in accuracy performance. The CNN used in our model is more robust against small movements of the face (e.g., slight facial position changes due to phone holding) but its performance was affected by more significant movement such as talking or head rotations.

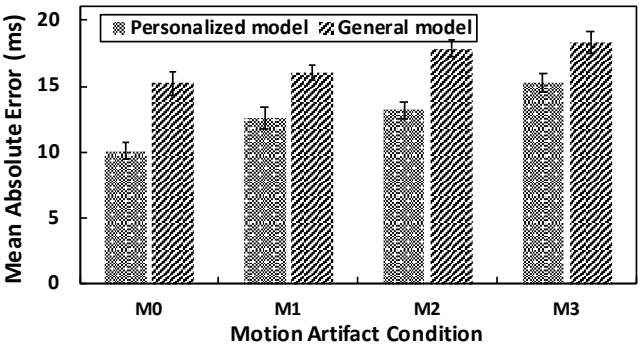
The peak position errors in both tables show that the Phase-1 model well executes the task of extracting the sub-frame that contains the peak of a heartbeat cycle. The MAE was maintained below a single frame for light conditions and two frames for motion artifacts. This suggests that if *VitaMon* needs to consider a maximum of 5 frames to calculate the exact time of the peak in the second phase.

## 6.2 Inter-beat Interval

Next, we examine the accuracy of *VitaMon* to predict the inter-beat intervals (IBIs). (Note: IBI measures the distance between two




(a) Peak detection MAE for different light conditions

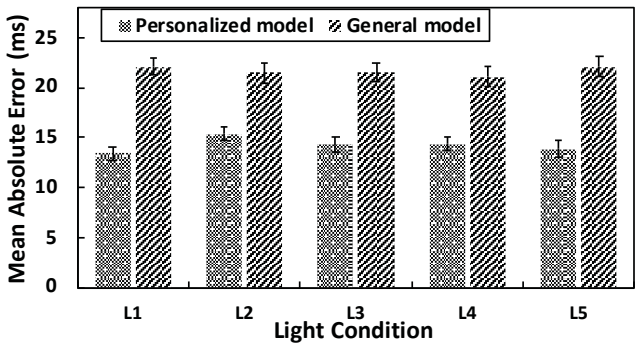



(b) Peak detection MAE for different motion artifacts

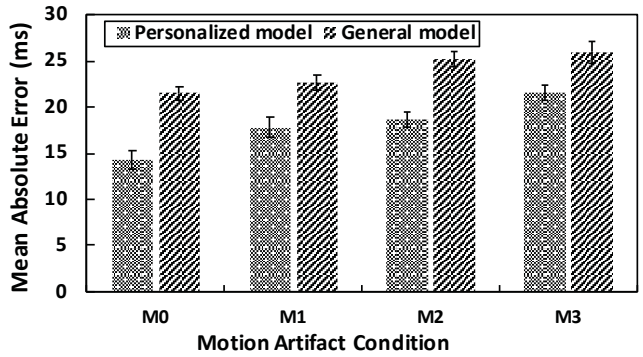
**Figure 9: Mean absolute error of peak detection in *VitaMon* .**

R-peaks in an ECG and is used to capture disorders such as arrhythmia.) To compute an accurate IBI, we utilize the full *VitaMon* system, including the Phase 2 model for capturing fine-grain heartbeat occurrence times. As in heart rate evaluations, we test the performance of *VitaMon* for different lighting conditions and motion artifacts with the personalized and global models.

An accurate IBI measurement requires the precise detection of peak times in heartbeat cycles. For this, we first measure the MAE for estimated peak times. Figure 9 presents the results for different




(a) IBI MAE for different light conditions




(b) IBI MAE for different motion artifacts

**Figure 10: Mean absolute error of inter-beat interval (IBI) measurements.**

light conditions and motion artifacts. The results show that *VitaMon* estimates the peak times with the errors of around only 10ms for the personal model and 15 ms for the global model. Also, accuracy is minimally affected by different illumination levels and motion artifacts. Assuming an 80 bpm heartbeat, a 10 msec error translates to an error of only 1.3% on the time-scale. Even when using the general model and also when introducing different motion artifacts, we observe errors less than 17 msec. This suggests that *VitaMon*'s Phase-2 model estimates the heartbeat cycle peak (ECG R-peak) with very high accuracy; its underlying CNN model well captures the correlation between a face image (that include multiple PPG data points at different facial areas) and the actual peak time.

We then evaluate the IBI estimation accuracy of *VitaMon* . Figure 10 shows the results. The MAEs for the IBI, in all cases, are below 22 msec, and for the personalized model the errors are as low as 12 msec. This is expected as the accurate peak detection contributes to the accurate calculation of the IBI. There is a slight impact on the performance as motion artifacts are introduced, but this increase can be considered minimal considering their applicability in many applications that involve users' mobility, especially in mobile context.

We also evaluate the performance of *VitaMon* on three groups of participants with different skin tones. The results are summarized in Table 5. Note that melanin, the pigment that accounts for the color of human skin, has a high absorption coefficient compared to

| | | Skin Tone Group | | |
|---|---|---|---|---|
| Metric | Model | G0 | G1 | G2 |
| HR MAE | General | 0.58 | 1.90 | 0.71 |
| (bpm) | Personalized | 0.32 | 1.55 | 0.53 |
| IBI MAE | General | 20.71 | 25.43 | 20.92 |
| (ms) | Personalized | 13.17 | 19.45 | 12.09 |

**Table 5: Evaluation per skin tone group under stationary condition. G0: light yellow skin tone, N = 22; G1: dark brown skin tone, N = 6; G2: white skin tone, N = 2**

hemoglobin's in the wavelength range of visible light. Hence, more melanin in skin or darker skin tone would attenuate the strength of optical signal of blood volume pulse. Compared to the other two groups, estimation on the group of participants with dark skin tone has significantly higher error, in terms of both heart rate and heartbeat interval measurement. On the other hand, Table 5 shows a similar evaluation results of *VitaMon* on participants with light yellow skin tone and participants with white skin tone, sample size of the latter group is small (N = 2) though.

## 6.3 HRV Features

Next, we evaluate how accurately *VitaMon* calculates various HRV features using the detected peak times. For HRV evaluation, we extract a list of standard features in the time-domain, geometric Poincare plot, and frequency-domain widely used for clinical purposes [3, 7, 35]. Specifically, *RMSSD* is the square root of the mean of the squares of successive differences between adjacent intervals, *SDNN* is the standard deviation of intervals, *SDSD* is the standard deviation of the successive differences between adjacent intervals, *NN*50 shows the number of pairs of successive intervals that differ by more than 50 msec, and *pNN*50 represents the proportion of NN50 divided by the total number of intervals. These metrics are features included in the time-domain. For the geometric Poincare plot features, *SD*1 shows the length of the longitudinal line in the Poincare plot of the intervals, and *SD*2 is the length of the transverse line in the Poincare plot of intervals. Lastly for features in the frequency domain, $LFnu$ shows the normalized spectral power in the low-frequency band from 0.04 to 0.15 Hz, and $HFnu$ is the normalized spectral power in the high-frequency band from 0.15 to 0.4 Hz. We point interested readers to [2, 37] for more details on these metrics.

Table 6 presents a comparison between the *VitaMon* -estimated features and ECG-driven features (used as the ground truth). From the correlation coefficients, we can see that for five of the nine features (i.e., RMSSD, SDNN, MRRI, SD1 and SD2), *VitaMon* achieves a very high correlation with the ground truth. For the two frequency domain features ($LFnu$ and $HFnu$ ), the correlations were 0.71 which are lower than the correlations of other time-domain features. This is because the frequency-domain features represent the trend in interval series and require accurate estimation of multiple continuous data points (intervals) to capture. In particular, LF band covers 0.04-0.15Hz or 7-15 second rhythm of interval series.

| | | HRV features | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Statistic | Source | RMSSD | SDNN | MRRI | NN50 | PNN50 | SD1 | SD2 | LFnu | HFnu |
| Mean | ECG | 111.65 | 87.32 | 751.92 | 9.90 | 13.50 | 80.39 | 96.01 | 25.67 | 74.40 |
| Mean | VitaMon | 114.61 | 89.30 | 749.10 | 33.54 | 46.10 | 79.62 | 95.70 | 33.74 | 65.74 |
| Standard Deviation | ECG | 70.58 | 45.07 | 68.52 | 9.27 | 12.38 | 50.12 | 42.84 | 30.45 | 30.45 |
| Standard Deviation | VitaMon | 54.43 | 38.58 | 68.37 | 10.19 | 12.05 | 39.09 | 39.42 | 21.92 | 21.70 |
| Correlation Coefficient | | 0.9817 | 0.9776 | 0.9943 | 0.4697 | 0.4317 | 0.9717 | 0.9710 | 0.72 | 0.72 |

*RMSSD*: The square root of the mean of the squares of successive differences between adjacent intervals.
*SDNN*: The standard deviation of intervals.
*MRRI*: The mean of R-R intervals.
*SDSD*: The standard deviation of the successive differences between adjacent intervals.
*NN*50: The number of pairs of successive intervals that differ by more than 50 msec.
*pNN*50: The proportion of NN50 divided by the total number of intervals.
*SD*1: The length of the longitudinal line in the Poincare plot of the intervals.
*SD*2: The length of the transverse line in the Poincare plot of intervals.
*LFnu*: The normalized spectral power in the low-frequency band from 0.04 to 0.15 Hz.
*HFnu*: The normalized spectral power in the high-frequency band from 0.15 to 0.4 Hz.

**Table 6: HRV monitoring performance of the general model: Average HRV features extracted from ECG reference signal and VitaMon estimation under stationary condition.**

However, the correlations are still high; our evaluation for stress detection (estimated by the ratio between *LFnu* and *HFnu*) in Section 7 shows that the accuracy of stress detection using the *LFnu* and *HFnu* features estimated by *VitaMon* was comparable with the same features extracted from the ECG reference signal. The errors for *NN*50 and *pNN*50 were high; these features are calculated based on the difference between two heartbeat intervals and the error of *VitaMon* 's IBI estimation could be doubled while there is a clear binary threshold of 50ms for evaluation.

The issues observed from the global model in Table 6 are alleviated in Table 7, where we plot the results for a personal model. This is so due to the fact that a personalized model will show fewer variations with higher peak detection accuracy as its underlying CNN models better captures the relationships between actual ECG signal peaks and the front camera images. While we omit the results for the case with different motion artifacts, similar trends were observed with other earlier evaluations.

### 6.4 Evaluation for Samples Collected from Real-world Use Cases

We also evaluate *VitaMon* on the data collected while driving and chatting in a coffee shop described in Section 5. Table 8 shows that *VitaMon* can measure heart rate with the errors of 1-2bpm using our Phase-1 model. The errors for the peak detection and inter-beat interval are higher than in the lab experiment, however, the errors remain low; for instance, the inter-beat interval errors remain under 23 ms for the personalized model. We attribute the increment of the errors to the different light conditions and motion artifact that are not captured in our training data; for instance, passengers' mobile phones were shaken when the car accelerated. We believe we can further improve the accuracy of our models in various ways. For instance, we can train our model with a more diverse set of data collected in real-life situations. Also, it is possible to use the phone's accelerometer data to filter out the segment of unstable video recording caused by the hand's motion artifact.

## 7 *VITAMON* APPLICATIONS

*VitaMon* can be applied to various useful applications. Online education is one example of where *VitaMon* can play an important role. As a student participates in the education programs, we can continuously monitor their engagement and stress levels using a face-facing camera, which are features that are known to be heavily correlated with HRV [10, 11, 21, 43, 51].

### 7.1 User Study

To study the feasibility of applying *VitaMon* to capture the cardiovascular responses to such psychological distress situations, we conduct a small user study that involves 12 participants (age from 26 to 35). This user study includes an arithmetic stress test session and a baseline session. All 12 participants participated in both of these sessions. In the arithmetic stress test session, we follow the validated experiment procedure described in [29, 34, 60]. While the users were facing the front camera on a smartphone, we verbally delivered questions with simple arithmetic operations (subtract 13 from 1022 as fast and accurately as possible), and their responses from mental calculation was delivered back to us verbally as well. Upon responding with an incorrect answer, the participants re-started the process from 1022, based on verbal feedback indicating to restart the calculation. Note that the arithmetic test was used as a tool to induce a psychological distress situation and the subjective stress level under the test may vary among participants. However, we did not collect the self-report stress level as the main purpose of the test is not to classify participants' stress level, but to study the cardiovascular responses measured by *VitaMon* as compared to the features from ECG reference signal. The baseline session was designed so that the participants stay still (sitting) while soothing classical music was

| | | HRV features | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Statistic | Source | RMSSD | SDNN | MRRI | NN50 | PNN50 | SD1 | SD2 | LFnu | HFnu |
| Mean | ECG | 112.20 | 88.51 | 751.00 | 9.88 | 13.46 | 80.36 | 95.39 | 25.97 | 74.10 |
| Mean | VitaMon | 114.51 | 89.47 | 750.15 | 16.46 | 23.30 | 80.55 | 95.77 | 30.66 | 68.81 |
| Standard Deviation | ECG | 69.94 | 44.82 | 68.52 | 9.33 | 12.46 | 49.66 | 42.79 | 30.54 | 30.54 |
| Standard Deviation | VitaMon | 61.29 | 41.49 | 68.38 | 9.55 | 13.02 | 43.97 | 4.90 | 24.71 | 24.50 |
| Correlation Coefficient | | 0.9879 | 0.9836 | 0.9855 | 0.7948 | 0.7394 | 0.9861 | 0.9830 | 0.8134 | 0.8139 |

Please refer to Table 6 for an explanation of each feature.

**Table 7: HRV monitoring performance of the personal model: Average HRV Features extracted from ECG reference signal and VitaMon estimation under stationary condition.**

| | Personalized model | | General model | |
|---|---|---|---|---|
| Metric | R1 | R2 | R1 | R2 |
| HR MAE (bpm) | 1.25 | 1.00 | 2.00 | 2.00 |
| Peak MAE (frame) | 1.18 | 1.29 | 1.52 | 1.48 |
| Peak MAE (ms) | 16.97 | 15.25 | 19.10 | 17.55 |
| IBI MAE (ms) | 22.57 | 19.98 | 25.40 | 22.99 |

**Table 8: Evaluation on data collected from real-world scenarios: (R1) passenger in a driving car and (R2) in coffee shop with dim light - 40lux.**

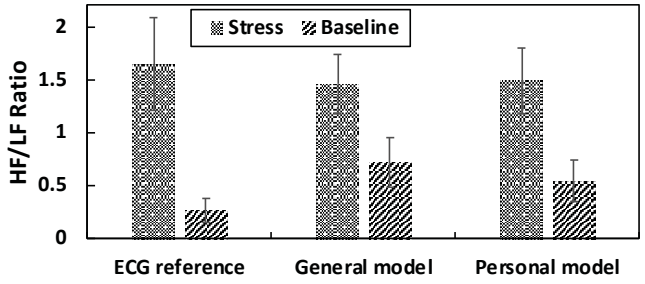



**Figure 11: HF/LF ratio HRV feature to distinguish stress and baseline condition.**

played. We used the Lenovo Phab 2 smartphone for data collection and kept each session for five minutes each. Upon the beginning of the first session, three minutes were given to the participants to minimize the effect of the previous session. The two sessions were separated by a six-minute break, and the session order was randomized [29].

## 7.2 Data Analysis

To examine the cardiovascular responses to the stress stimulation, we used a well-adopted previous method proven to be effective for stress detection [9]. The analysis of HRV was carried out using the low frequency (LF;0.04-0.15 Hz) and high frequency (HF; 0.15-0.40 Hz) bands, which reflect the sympathetic activity with vagal modulation, and parasympathetic activity, respectively.

Figure 11 plots the comparison of the ratio of HF and LF for the ECG-based baseline, *VitaMon* with the global model and, *VitaMon* with the personalized model when participating in the two different sessions. Results suggest that indeed when the participant is involved in the arithmetic stress test session, the ratio of HF over LF shows a noticeably high value compared to the case when the participant is in the baseline session. The results are consistent with prior studies [10, 11] showing that under mental stress condition, the HF spectral power increases while LF power decreases. The figure also serves as an indicator for suggesting that *VitaMon* can be a useful involuntary sensing tool for measuring stress. In Table 9 we present additional details on the observations made for each study participant. The results suggest that both types of models can effectively be used for a real-world application to detect stress levels. We also emphasize that when observing the HRV features themselves, the frequency-band features, $LFnu$ and $HFnu$ did not show a significantly high correlation with the ground truth. However, when utilizing these features as application-specific features, even such features can be considered useful for the target purpose.

# 8 DISCUSSION

In this section, we discuss some of the limitations of *VitaMon* and present our future work plans.

## 8.1 Effect of Skin Tone & Make-up on *VitaMon*

*VitaMon* uses a camera-based PPG method to extract the subtle variation of skin color caused the changing blood volume due to heartbeats. However, the degree of color variation seen on the face by the camera also depends on facial features such as the color and intensity of the skin pigments and the amount and type of make-up used etc. In particular, dark skin pigments or heavy make-up have a high light absorption coefficient, which would result in a weaker type of pulse signal being observed in the facial video. We plan to extend our tests of *VitaMon* across a larger population segment in future work.

| | | Subject | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Source | Condition | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 |
| ECG Reference | Stress | 1.05 | 2.76 | 2.90 | 4.50 | 0.03 | 0.03 | 0.03 | 0.04 | 1.90 | 3.23 | 1.22 | 2.02 |
| ECG Reference | Baseline | 0.02 | 0.04 | 0.04 | 0.04 | 0.03 | 0.04 | 0.57 | 0.08 | 0.03 | 1.28 | 0.49 | 0.52 |
| Personal Model | Stress | 1.62 | 1.47 | 1.69 | 2.23 | 0.03 | 0.49 | 0.43 | 0.27 | 1.86 | 3.89 | 1.91 | 2.06 |
| Personal Model | Baseline | 0.03 | 1.23 | 0.05 | 0.13 | 0.04 | 0.05 | 0.66 | 0.13 | 0.06 | 1.88 | 0.53 | 1.68 |
| General Model | Stress | 1.54 | 1.59 | 1.23 | 2.21 | 0.08 | 0.68 | 0.38 | 0.32 | 1.88 | 2.61 | 1.91 | 3.16 |
| General Model | Baseline | 0.10 | 0.62 | 0.13 | 0.96 | 0.07 | 0.16 | 0.71 | 0.09 | 0.14 | 1.36 | 1.53 | 2.70 |

**Table 9: Average HRV Features extracted from ECG reference signal and VitaMon estimation (personal model) under stationary condition.**

### 8.2 Integrating *VitaMon* With Built-in Camera Optimisations

Modern smart phone cameras perform a number of automatic image corrections to improve the quality of the images taken as perceived by a human user. For example, the camera might automatically sharpen or increase the contrast of the image or even brighten the image if the ambient light is too low. In addition, many smart phone cameras automatically perform color filtering to increase the vividness of the photos and videos. In this paper, we did not investigate how *VitaMon* would operate in situations where the camera software was automatically manipulating the images using in-built algorithms.

### 8.3 Measuring Other Related Cardiovascular Signals

In addition to detecting HRV, *VitaMon* may also be able to detect other related physiological signals. In particular, many prior studies show that, a person's respiration rate can be extracted directly from the continuous IBI series, which we are already collecting in *VitaMon*. Furthermore, given that the pulse speed is known to be inversely related to the blood pressure, we can utilize the pulse propagation delay utilized in this work for continuous blood pressure monitoring. With the right extensions to our model, *VitaMon* could accurately detect these signals using the same input data. We plan to investigate this in the future.

### 8.4 Limitations & Future Work

The user study was conducted mainly with student volunteers in two countries. It is possible that a more diverse user pool would show very different results. In the future, we plan to improve *VitaMon* by 1) extending it to detect other physiological signals, 2) improving its performance by integrating a simple yet powerful training step – where a user can quickly provide facial data that is added to a pre-trained general model to create a much better performing semi-personalised model. Finally, 3) we plan to integrate *VitaMon* into a student life-logging app and deploy it more generally across a larger audience.

## 9 CONCLUSION

We present *VitaMon*, a mobile sensing system for daily HRV monitoring using a commodity smartphone's front camera. We first present our two key insights in designing *VitaMon*: a human face contains *multiple* cardiovascular pulse signals with *different* phase shift. Then, we build a CNN-based technique to extract both spatial and temporal information of the video to reconstruct a pulse waveform signal that is optimized for detecting the exact time of heartbeat cycle peak occurrences, from which inter-beat intervals (IBIs) and HRV features can be calculated. We evaluated *VitaMon* with a dataset collected from 30 participants under various conditions involving different light intensity levels and motion artifacts. Our results show that, with 15 fps video inputs (66.67 ms time resolution), *VitaMon* can measure IBI with an average error of 14.26 ms and 21.65 ms using personal and general models, respectively. Both time- and frequency-domain HRV features extracted from the IBI measurements show a high linear relationship with the reference signal.

## 10 ACKNOWLEDGMENTS

We thank the anonymous shepherd and reviewers for their insightful comments. This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2018R1C1B6003869) and the National Research Foundation of Korea (NRF) grant (No. 2019R1C1C1006088). Rajesh K. Balan, JeongGil Ko, and Youngki Lee are co-corresponding authors of this work.

## REFERENCES

[1] 2016. External Carotid Artery. Available at: https://www.stepwards.com/?page_id=5802. (2016).
[2] U Rajendra Acharya, K Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S Suri. 2006. Heart rate variability: a review. *Medical and biological engineering and computing* 44, 12 (2006), 1031–1051.
[3] Nutan D Ahuja, Amit K Agarwal, Ninad M Mahajan, Naresh H Mehta, and Hatim N Kapadia. 2003. GSR and HRV: its application in clinical diagnosis. In *16th IEEE Symposium Computer-Based Medical Systems, 2003. Proceedings.* IEEE, 279–283.
[4] R. Almeida, S. Gouveia, A. P. Rocha, E. Pueyo, J. P. Martinez, and P. Laguna. 2006. QT variability and HRV interactions in ECG: quantification and reliability. *IEEE Transactions on Biomedical Engineering* 53, 7 (July 2006), 1317–1329. https://doi.org/10.1109/TBME.2006.873682
[5] Christoph Hoog Antink, Hanno Gao, Christoph Brüser, and Steffen Leonhardt. 2015. Beat-to-beat heart rate estimation fusing multimodal video and sensor data. *Biomedical optics express* 6, 8 (2015), 2895–2907.

[6] Mordechai Azaria and David Hertz. 1984. Time delay estimation by generalized cross correlation methods. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 32, 2 (1984), 280–285.
[7] George E Billman, Heikki V Huikuri, Jerzy Sacha, and Karin Trimmel. 2015. An introduction to heart rate variability: methodological considerations and clinical applications. *Frontiers in physiology* 6 (2015), 55.
[8] M Bolanos, H Nazeran, and E Haltiwanger. 2006. Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 4289–4294.
[9] AJMM Camm, Marek Malik, JTGB Bigger, Günter Breithardt, Sergio Cerutti, R Cohen, Philippe Coumel, E Fallen, H Kennedy, RE Kleiger, et al. 1996. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation* 93, 5 (1996), 1043–1065.
[10] Rossana Castaldo, Paolo Melillo, Umberto Bracale, M Caserta, Maria Triassi, and Leandro Pecchia. 2015. Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomedical Signal Processing and Control* 18 (2015), 370–377.
[11] Mariana N Castro, Daniel E Vigo, Elvina M Chu, Rodolfo D Fahrer, Delfina de Achával, Elsa Y Costanzo, Ramón C Leiguarda, Martín Nogués, Daniel P Cardinali, and Salvador M Guinjoan. 2009. Heart rate variability response to mental arithmetic stress is abnormal in first-degree relatives of individuals with schizophrenia. *Schizophrenia research* 109, 1-3 (2009), 134–140.
[12] Amer Abdulmahdi Chlaihawi, Binu Baby Narakathu, Sepehr Emamian, Bradley J Bazuin, and Massood Z Atashbar. 2018. Development of printed and flexible dry ECG electrodes. *Sensing and bio-sensing research* 20 (2018), 9–15.
[13] B. Barkat D. Boutana, M. Benidir. 2011. Segmentation and identification of some pathological phonocardiogram signals using time-frequency analysis. *IET Signal Processing* 5 (September 2011), 527–537(10). Issue 6. https://digital-library.theiet.org/content/journals/10.1049/iet-spr.2010.0013
[14] Maria I Davila, Gregory F Lewis, and Stephen W Porges. 2017. The Physiocam: a novel non-contact sensor to Measure heart rate Variability in clinical and Field applications. *Frontiers in public health* 5 (2017), 300.
[15] Elias Ebrahimzadeh, Mohammad Pooyan, and Ahmad Bijar. 2014. A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. *PloS one* 9, 2 (2014), e81896.
[16] LA Geddes, MH Voelz, CF Babbs, JD Bourland, and WA Tacker. 1981. Pulse transit time as an indicator of arterial blood pressure. *psychophysiology* 18, 1 (1981), 71–74.
[17] Przemyslaw Guzik and Marek Malik. 2016. ECG by mobile technologies. *Journal of Electrocardiology* 49, 6 (2016), 894 – 901. https://doi.org/10.1016/j.jelectrocard.2016.07.030
[18] Song Han, Huizi Mao, and William J Dally. 2015. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149* (2015).
[19] James AJ Heathers. 2013. Smartphone-enabled pulse rate variability: an alternative methodology for the collection of heart rate variability in psychophysiological research. *International Journal of Psychophysiology* 89, 3 (2013), 297–304.
[20] Alrick B Hertzman. 1937. Observations on the finger volume pulse recorded photoelectrically. *Am. J. Physiol.* 119 (1937), 334–335.
[21] Nis Hjortskov, Dag Rissén, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Søgaard. 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology* 92, 1-2 (2004), 84–89.
[22] Loc N Huynh, Youngki Lee, and Rajesh Krishna Balan. 2017. Deepmon: Mobile gpu-based deep learning framework for continuous vision applications. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 82–95.
[23] JR Jago and A Murray. 1988. Repeatability of peripheral pulse measurements on ears, fingers and toes using photoelectric plethysmography. *Clinical Physics and Physiological Measurement* 9, 4 (1988), 319.
[24] Vala Jeyhani, Shadi Mahdiani, Mikko Peltokangas, and Antti Vehkaoja. 2015. Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 5952–5955.
[25] James A Johnstone, Paul A Ford, Gerwyn Hughes, Tim Watson, and Andrew T Garrett. 2012. BioHarness multivariable monitoring device: part. I: validity. *Journal of sports science & medicine* 11, 3 (2012), 400.
[26] James A Johnstone, Paul A Ford, Gerwyn Hughes, Tim Watson, and Andrew T Garrett. 2012. BioharnessâĎć multivariable monitoring device: part. II: reliability. *Journal of sports science & medicine* 11, 3 (2012), 409.
[27] Seungwoo Kang, Sungjun Kwon, Chungkuk Yoo, Sangwon Seo, Kwangsuk Park, Junehwa Song, and Youngki Lee. 2014. Sinabro: Opportunistic and Unobtrusive Mobile Electrocardiogram Monitoring System. In *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (HotMobile '14)*. ACM, New York, NY, USA, Article 11, 6 pages. https://doi.org/10.1145/2565585.2565605
[28] J-H Kim, R Roberge, JB Powell, AB Shafer, and W Jon Williams. 2013. Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the Zephyr BioHarnessâĎć. *International journal of sports medicine* 34, 6 (2013), 497.
[29] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. 1993. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 1-2 (1993), 76–81.
[30] Charles Knapp and Glifford Carter. 1976. The generalized correlation method for estimation of time delay. *IEEE transactions on acoustics, speech, and signal processing* 24, 4 (1976), 320–327.
[31] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal. 2015. DistancePPG: Robust non-contact vital signs monitoring using a camera. *Biomedical optics express* 6, 5 (2015), 1565–1588.
[32] Sungjun Kwon, Hyunseok Kim, and Kwang Suk Park. 2012. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2174–2177.
[33] Sungjun Kwon, Jeehoon Kim, Dongseok Lee, and Kwangsuk Park. 2015. ROI analysis for remote photoplethysmography on facial video. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4938–4941.
[34] Wolfgang Linden. 1991. What do arithmetic stress tests measure? Protocol variations and cardiovascular responses. *Psychophysiology* 28, 1 (1991), 91–102.
[35] Federico Lombardi. 2002. Clinical implications of present physiological understanding of HRV components. *Cardiac electrophysiology review* 6, 3 (2002), 245–249.
[36] S. Mahdiani, V. Jeyhani, M. Peltokangas, and A. Vehkaoja. 2015. Is 50 Hz high enough ECG sampling frequency for accurate HRV analysis?. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 5948–5951. https://doi.org/10.1109/EMBC.2015.7319746
[37] Marek Malik and A John Camm. 1990. Heart rate variability. *Clinical cardiology* 13, 8 (1990), 570–576.
[38] Kenta Matsumura and Takehiro Yamakoshi. 2013. iPhysioMeter: a new approach for measuring heart rate and normalized pulse volume using only a smartphone. *Behavior research methods* 45, 4 (2013), 1272–1278.
[39] Daniel McDuff, Sarah Gontarek, and Rosalind Picard. 2014. Remote measurement of cognitive stress via heart rate variability. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2957–2960.
[40] Goris Nazari, Joy C Macdermid, Kathryn E Sinden R Kin, Julie Richardson, and Ada Tang. 2018. Reliability of Zephyr bioharness and Fitbit charge measures of heart rate and activity at rest, during the modified Canadian aerobic fitness test and recovery. *The Journal of Strength & Conditioning Research* (2018).
[41] D. Nepi, A. Sbrollini, A. Agostinelli, E. Maranesi, M. Morettini, F. Di Nardo, S. Fioretti, P. Pierleoni, L. Pernini, S. Valenti, and L. Burattini. 2016. Validation of the heart-rate signal provided by the Zephyr bioharness 3.0. In *2016 Computing in Cardiology Conference (CinC)*. 361–364. https://doi.org/10.23919/CIC.2016.7868754
[42] The National Optical Astronomy Observatory. 2015. Recommended Light Levels (Illuminance) for Outdoor and Indoor Venues. Available at: https://www.noao.edu/education/QLTkit/ACTIVITY_Documents/Safety/LightLevels_outdoor+indoor.pdf. (2015).
[43] Gewnhi Park, Jay J Van Bavel, Michael W Vasey, and Julian F Thayer. 2013. Cardiac vagal tone predicts attentional engagement to and disengagement from fearful faces. *Emotion* 13, 4 (2013), 645.
[44] Leandro Pecchia, Paolo Melillo, and Marcello Bracale. 2011. Remote health monitoring of heart failure with data mining via CART method on HRV features. *IEEE Transactions on Biomedical Engineering* 58, 3 (2011), 800–804.
[45] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. 2010. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express* 18, 10 (2010), 10762–10774.
[46] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. 2011. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering* 58, 1 (2011), 7–11.
[47] Angel Melchor Rodríguez and J Ramos-Castro. 2018. Video pulse rate variability analysis in stationary and motion conditions. *Biomedical engineering online* 17, 1 (2018), 11.
[48] Alex Sagie, Martin G Larson, Robert J Goldberg, James R Bengtson, and Daniel Levy. 1992. An improved method for adjusting the QT interval for heart rate (the Framingham Heart Study). *The American journal of cardiology* 70, 7 (1992), 797–801.
[49] J Michael Schmidt. 2016. Heart rate variability for the early detection of delayed cerebral ischemia. *Journal of Clinical Neurophysiology* 33, 3 (2016), 268–274.
[50] Nandakumar Selvaraj, Ashok Jaryal, Jayashree Santhosh, Kishore K Deepak, and Sneh Anand. 2008. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *Journal of medical engineering & technology* 32, 6 (2008), 479–484.

[51] Piia Seppälä, Saija Mauno, Marja-Liisa Kinnunen, Taru Feldt, Tanja Juuti, Asko Tolvanen, and Heikki Rusko. 2012. Is work engagement related to healthy cardiac autonomic activity? Evidence from a field study among Finnish women workers. *The Journal of Positive Psychology* 7, 2 (2012), 95–106.

[52] Ahamed Seyd, Paul K Joseph, and Jeevamma Jacob. 2012. Automated diagnosis of diabetes using heart rate variability signals. *Journal of medical systems* 36, 3 (2012), 1935–1941.

[53] Emilia Sforza, Vincent Pichot, Katerina Cervena, Jean Claude Barthélémy, and Frederic Roche. 2007. Cardiac variability and heart-rate increment as a marker of sleep fragmentation in patients with a sleep disorder: a preliminary study. *Sleep* 30, 1 (2007), 43–51.

[54] Robin P Smith, Jérôme Argod, Jean-Louis Pépin, and Patrick A Lévy. 1999. Pulse transit time: an appraisal of potential clinical applications. *Thorax* 54, 5 (1999), 452–457.

[55] Phyllis K Stein and Yachuan Pu. 2012. Heart rate variability, sleep and sleep disorders. *Sleep medicine reviews* 16, 1 (2012), 47–66.

[56] Yu Sun, Sijung Hu, Vicente Azorin-Peris, Roy Kalawsky, and Stephen E Greenwald. 2012. Noncontact imaging photoplethysmography to effectively access pulse rate variability. *Journal of biomedical optics* 18, 6 (2012), 061205.

[57] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826.

[58] XF Teng and Yuan-Ting Zhang. 2004. The effect of contacting force on photoplethysmographic signals. *Physiological measurement* 25, 5 (2004), 1323.

[59] TensorFlow. 2019. TensorFlow Lite on GPU. Available at: https://www.tensorflow.org/lite/performance/gpu_advanced. (2019).

[60] Robert L Trestman, Emil F Coccaro, David Bernstein, Timothy Lawrence, Steven M Gabriel, Thomas B Horvath, and Larry J Siever. 1991. Cortisol responses to mental arithmetic in acute and remitted depression. *Biological psychiatry* 29, 10 (1991), 1051–1054.

[61] Brooke D Vergales, Santina A Zanelli, Julie A Matsumoto, Howard P Goodkin, Douglas E Lake, J Randall Moorman, and Karen D Fairchild. 2014. Depressed heart rate variability is associated with abnormal EEG, MRI, and death in neonates with hypoxic ischemic encephalopathy. *American journal of perinatology* 31, 10 (2014), 855–862.

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

[63] C. Wang, S. Yu, Y. Lin, and Y. Lin. 2016. Fatigue detection system based on indirect-contact ECG measurement. In *2016 International Conference on Advanced Robotics and Intelligent Systems (ARIS)*. 1–1. https://doi.org/10.1109/ARIS.2016.7886623

[64] Matthias Weippert, Mohit Kumar, Steffi Kreuzfeld, Dagmar Arndt, Annika Rieger, and Regina Stoll. 2010. Comparison of three mobile devices for measuring R–R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. *European Journal of Applied Physiology* 109, 4 (01 Jul 2010), 779–786. https://doi.org/10.1007/s00421-010-1415-9

[65] J. M. Zanetti and D. M. Salerno. 1991. Seismocardiography: a technique for recording precordial acceleration. (May 1991), 4–9. https://doi.org/10.1109/CBMS.1991.128936