TrueHeart: Continuous Authentication on Wrist-worn Wearables Using PPG-based Biometrics

Tianming Zhao*, Yan Wang*, Jian Liu†‡, Yingying Chen‡, Jerry Cheng§ and Jiadi Yu¶
*Temple University, Philadelphia, PA, USA 19122
†The University of Tennessee, Knoxville, TN, USA 37996
‡Rutgers University, New Brunswick, NJ, USA 08901
§New York Institute of Technology, New York, NY, USA 10023
¶Shanghai Jiao Tong University, Shanghai, P.R. China 200240
Email:{tum94362,y.wang}@temple.edu, jliu@utk.edu, yingche@scarletmail.rutgers.edu,
jcheng18@nyit.edu, jiadiyu@sjtu.edu.cn

***Abstract*—Traditional one-time user authentication processes might cause friction and unfavorable user experience in many widely-used applications. This is a severe problem in particular for security-sensitive facilities if an adversary could obtain unauthorized privileges after a user's initial login. Recently, continuous user authentication (CA) has shown its great potential by enabling seamless user authentication with few active participation. We devise a low-cost system exploiting a user's pulsatile signals from the photoplethysmography (PPG) sensor in commercial wrist-worn wearables for CA. Compared to existing approaches, our system requires zero user effort and is applicable to practical scenarios with non-clinical PPG measurements having motion artifacts (MA). We explore the uniqueness of the human cardiac system and design an MA filtering method to mitigate the impacts of daily activities. Furthermore, we identify general fiducial features and develop an adaptive classifier using the gradient boosting tree (GBT) method. As a result, our system can authenticate users continuously based on their cardiac characteristics so little training effort is required. Experiments with our wrist-worn PPG sensing platform on 20 participants under practical scenarios demonstrate that our system can achieve a high CA accuracy of over 90% and a low false detection rate of 4% in detecting random attacks.**

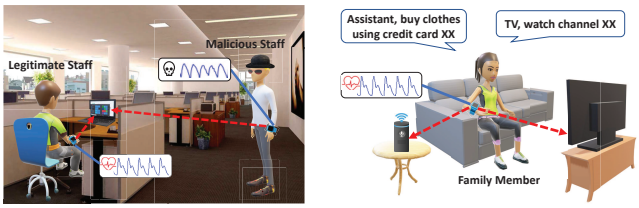
I. Introduction

Traditional user authentication methods rely on users' inputs, such as passwords and graphic patterns. However, these methods are known to be vulnerable to many attacks [1], [2]. Recently, multi-factor authentication (MFA) [3], [4] has been proposed to mitigate these threats by verifying two or more confidential information from independent sources. While many applications have adopted either one-factor or MFA, both of these two approaches use a one-time login process, which is not secure enough to authenticate users in the duration of certain applications. This is especially critical for a security-sensitive application, in which an adversary could obtain unauthorized privileges after a user's initial login. Therefore, a practical continuous user authentication (CA) solution that can periodically verify a user's identity without interruptions of the application usage is highly in demand [5].

Existing CA approaches usually focus on reducing or eliminating user involvement in the authentication process by leveraging users' unique behavioral patterns. For example, keystroke/mouse dynamics [6], [7] and gait patterns [8] have been used for user authentication since 2012. These approaches usually rely on momentary events and can only determine a user's identity by monitoring particular activities (e.g., typing, mouse-clicking, or walking). There are studies using cardiac signals (e.g., ECG [9], [10] and cardiac motion [11]) for CA. All these systems require dedicated sensors (e.g., ECG or Doppler radar sensors), which are costly and not readily available in commodity devices. Recently, researchers find that the photoplethysmography (PPG) sensor can also provide unique cardiac biometric information for user authentication [12]–[15]. However, these systems only focus on clinical scenarios, under which strong and stable PPG measurements are collected from the fingertips of static subjects.

Different from the existing works, we develop a low-cost CA system, TrueHeart, which can periodically verify the identity of a user via cardiac signals (i.e., PPG) from common wrist-worn wearable devices (e.g., smartwatches and fitness trackers). Under a working environment shown in Figure 1(a), TrueHeart can continuously determine whether a current staff operating a specific device (e.g., a smartphone or a laptop) is a legitimate user in a non-intrusive manner so that any time-sensitive tasks will not be interrupted. As a result, a user can continuously trade stocks, manage air traffic, or switch circuits. As a daily life example in Figure 1(b), each family member with a wearable device can be periodically authenticated by TrueHeart so that he/she can enjoy a seamless experience of accessing or switching between user-specific apps on the smart devices paired with TrueHeart. Therefore, each person can watch his/her own favorite channels in a smart TV or do online shopping via a voice assistant. The advantage of using PPG for CA is obvious as cardiac signals are unique and ever-present biometrics which are available without users' involvement. In addition, PPG requires physical contact to human skin and is usually hidden in the back of wearable devices. Therefore, PPG measurements are secure and difficult to counterfeit.

There are several challenges in performing CA using PPG measurements from wearable devices. First, *in contrast to ECG signals* which is electrical and generated by heart activities, PPG signals capture blood volume changes by measuring reflected light from human skins. Therefore, PPG signals are



(a) CA in office scenarios (b) CA in living scenarios

Fig. 1. Two scenarios of continuous user authentication (CA) using TrueHeart.

relatively coarse-grained, noisy, and more susceptible to interference than ECG signals. Although initial works [13], [16] have shown that PPG measurements from fingertips contain unique features to be used for user authentication in clinical environments. However, these features are not persistent in the PPG signals collected from wearable devices in practice. Second, wrist-worn wearable devices are usually associated with a lot of hand or body movements from daily activities. These movements would result in various motion artifacts (MAs) which make cardiac signals in PPG measurements often unavailable in practice. Third, due to various types of imprecisions in PPG sensors in wearable devices and loose contacts between them and human skins, cardiac signals from PPG measurements could vary among days or even in the same day.

To address these challenges, we particularly investigate and determine *general fiducial features* that are not only persistent in various users' PPG measurements but also can capture unique characteristics of cardiac motions for CA. Additionally, we study the MAs of different types of body-movements (e.g., walking, moving forearm, and drinking water) in practical scenarios and categorize them into two types: *far-wrist* and *near-wrist*, based on the recoverability of cardiac signals with the MAs. We further develop effective *MA detection* and *MA mitigation/removal mechanisms* to identify the two type of MAs and choose to either recover the cardiac signals from weak MA impacts or remove the measurements containing strong MA impacts. These mechanisms ensure that our CA system can extract correct cardiac signals without the impact from MAs and perform CA accurately under practical scenarios. Moreover, our system adopts an *adaptive updating mechanism* to automatically accommodate the user's cardiac signal changes over time based on adaptive training of associated classifiers. The main contributions of our work are summarized as follows:

- We develop TrueHeart, the first low-cost CA system, that can authenticate users by using unique cardiac biometrics extracted from PPG sensors in wrist-worn wearable devices. Our system can be easily deployed in any PPG-enabled wearable devices (e.g., smartwatches).
- We extensively study characteristics of MAs under many practical scenarios and develop robust MA mitigation and removal mechanisms that can effectively identify different types of MAs with various intensities and eliminate MA impact accordingly.
- We identify general fiducial features that can capture the uniqueness of users' cardiac patterns to build an adaptive gradient boosting tree (GBT)-based classifier that can be robust to signal drifts in PPG, authenticate users, and defend against random attack effectively.
- We build a prototype of TrueHeart using commodity PPG sensors. Experimental results involving 20 participants demonstrate that TrueHeart can achieve a high average CA accuracy of over 90% while maintaining a low false detection rate of 4% when detecting random attacks.

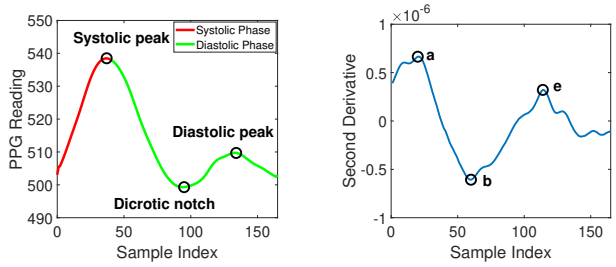
II. Related Work

Recent user authentication systems often use users' biometrics (e.g., behavioral or physiological information) to reduce user involvement and facilitate CA. Behavioral pattern is considered a distinct biometric that can make CA possible based on users' daily activities. For example, Mondol *et al.* [17] propose a user authentication system leveraging motion sensors in smartwatches to capture users' signatures in the air for authentication. Casale *et al.* [18] develop a wearable-based authentication system based on users' walk patterns. However, these approaches rely on users' involvement in specific activities in such a great deal to easily cause inconvenience.

Physiological-based biometrics (e.g., cardiac and respiratory motions) are popularly used for building CA systems because they can be obtained without users' active participation. For instance, Lin *et al.* [11] propose a CA system, Cardiac Scan, which utilizes DC-coupled continuous-wave radar to capture distinct heart motions in the user identification process. Rahman *et al.* [19] develop a method that uses the Doppler radar to identify users based on their respiratory motions. Although these systems provide a sound foundation for CA using wireless technology, they use dedicated devices that might not be available for users yet. Recently advanced sensing technologies enable unobtrusive and continuous user authentication based on unique cardiac biometrics captured by electrocardiogram (ECG) sensors [20], [21]. While mostly available under clinical environments, these systems require users to wear electrodes at various locations. This again turns out to be inconvenient for the uses in practice.

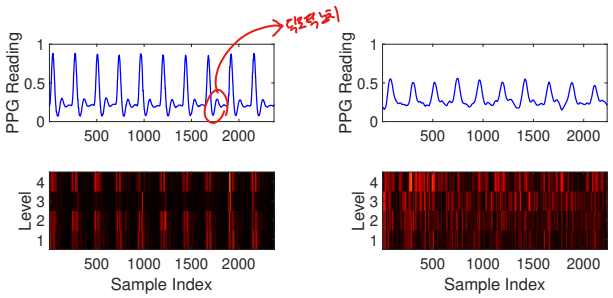
Unlike ECG, PPG is widely used in commodity wearable devices such as smartwatches and fitness trackers. Some initial studies have explored PPG-based authentications. For example, fiducial features [12], [13] have been discovered to capture unique characteristics in human cardiac systems so they can facilitate user authentication processes. Recently, non-fiducial features (i.e., discrete wavelet transform (DWT) coefficients) of PPG signals are proposed to build CA systems [14], [15]. However, all of the aforementioned studies collect PPG measurements from users' fingertips thus require users to wear dedicated PPG sensors and keep motionless. These requirements are different to meet in reality.

Different from the existing work, we build the first low-cost PPG-based system that can perform CA in practical scenarios with various body movements by leveraging PPG sensors in commodity wrist-worn devices. We identify general fiducial features that can capture distinct cardiac biometrics of diverse PPG measurements collected from users' wrist areas. In addition, we extensively study the impact of motions with



(a) Raw PPG measurements (b) Second derivative of raw PPG measurements

Fig. 2. Illustration of the critical landmarks in raw PPG measurements and its second derivative.



(a) Fingertip PPG data and DWT coefficients (b) Wrist PPG data and DWT coefficients

Fig. 3. Example of PPG data from fingertip & wrist and their corresponding discrete wavelet transform.

different intensities and develop the MA removal method that can effectively remove MA and significantly improve the CA performance. Moreover, our system employs an adaptive user authentication method that can reduce the impact of system drifts and provide long-term PPG-based CA.

III. Approach Overview

A. Attack Model

In this paper, we assume that attackers cannot compromise users' wearable devices (i.e., gaining access to their memory storages for raw PPG measurements). Based on this, the possible attacks to our CA system are as follows:

Random attack. Attackers or their accomplices wear users' wearable devices and expect the PPG measurements captured can pass our PPG-based CA system. This random attack model is similar to the brute-force attack.

Synthesis attack. To launch this attack, attackers first need to obtain users' blood flow patterns through either medical records or vision-based technologies (e.g., remote photoplethysmography (rPPG) [22]). However, these patterns and the PPG measurements collected from users' wrist areas are different in collection approaches and conditions. In addition, the PPG signals are collected in an enclosed environment (between the back of wearable devices and skin contact areas) so that many critical measurement data (light absorption/reflection of human skin, light source intensity, etc.) As a result, synthesis attacks will not be easily launched.

B. Feasibility Study

Intuition of Using Wearable PPG for CA. Human cardiac systems have been studied and known to be distinct among people [23]. Along this direction, initial studies [13], [16] have shown that *fiducial features* derived from critical landmarks in the raw PPG measurements and their derivatives (i.e., the systolic/diastolic peaks, dicrotic notch, and points a/b/c in Figure 2) can be used as users' unique biometric information. However, these studies only analyze PPG data collected from clinical settings with quite strict requirements. Thus, how to design and realize a PPG-based CA system using wrist-worn devices in practices remains a challenging task.

Difference between Wrist-Worn PPG and Fingertip PPG. To illustrate such a difference, we collect PPG measurements from both fingertip and wrist areas of the same users simultaneously using our prototype PPG sensing platform. The top two panels of Figure 3 show that the PPG measurements from the wrist area are stable but with less detectable and critical landmarks than those from the fingertip area. This indicates that the existing fiducial-feature-based authentication approaches [13], [16] are not applicable directly to the PPG from wearable devices. We further generate non-fiducial feature for both PPG measurements using the Daubechies wavelet of order 4 (db4) with four levels of decomposition. The bottom two panels in Figure 3 show that the fingertip PPG readings have repetitive and stable DWT coefficients with respect to each heartbeat in four levels, whereas the wrist area PPG readings are embedded with many noisy and irregular DWT coefficients, which will significantly impact the performance of the non-fiducial-based PPG authentication work [14], [15]. Therefore, instead of adopting non-fiducial features, there is a need to explore more general fiducial features in the PPG signal from the wrist area for CA, which is explained at *PPG Feature Extraction and User Authentication* in Section IV.

다우베키스 웨이블릿 (Daubechies Wavelet)은 웨이블릿 변환 (wavelet transform)에서 가장 널리 사용되는 함수군 중 하나로, 신호 처리와 데이터 압축, 노이즈 제거 등에 매우 유용한 도구

Impact of Daily Activities. To better understand the impact of daily activities as motion artifacts (MAs), we categorize them into three types based on the different moving parts of human bodies involved: *far-wrist*, *near-wrist*, and *whole-body* activities. The far-wrist activities are the major arm movements without involving tendons and muscles of the wrist area. In contrast, the near-wrist activities are finger-level and/or wrist-level movements, which have direct impacts on blood volume changes in the wrist area and more significant impact on PPG measurements from wearable devices. The whole-body activities are associated with most of human body parts. We find that some whole-body activities of low intensity, such as leisure walking, do not have noticeable impacts on the PPG measurements as shown in Figure 4. More strenuous activities, such as running, would change PPG readings significantly. In this work, we focus on the static and moving scenarios involving far-wrist and near-wrist activities, which cover the main scenarios in CA. We present the detailed design of our system in the following sections.

C. System Overview

The architecture of our PPG-based continuous user authentication system is shown in Figure 5. The system collects PPG measurements constantly from users' wearable devices as the input. Due to hardware imperfection, the raw PPG measurements inevitably contain baseline drifts and high-frequency interferences. Therefore, our system first performs

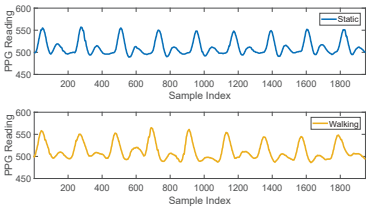


Fig. 4. The PPG measurements of the same people under static scenario and walking scenario.

Noise Reduction using Filtering to reduce such impacts. A band-pass filter is used to extract pulsatile components in PPG measurements. After filtering, the system conducts *Pulse Segmentation* to determine the PPG segment that is likely to contain a complete cardiac cycle. The insight is that each cardiac cycle should include a *systolic peak*, which could be identified in the PPG measurement during typical diastole and systole phases.

Next, we design *Motion Artifact (MA) Filtering* to remove MAs caused by daily physical activities. In PPG measurements, MAs arise from tissue deformations and local blood flow changes in the wrist area. While pulsatile signals are repetitive in PPG measurements, most MAs have burst PPG waveforms. We calculate statistical measures, such as kurtosis, skewness, and standard deviation, in pulse waveforms and MA signals to determine whether a PPG segment contains a pulse or an MA in the *MA detection* process. If MAs are detected, our system performs *MA Classification* to further decide whether they are from far-wrist activities or near-wrist activities. In general, near-wrist activities result in long-duration and strong and unrecoverable effects on PPG measurements, while far-wrist activities have small and recoverable impacts. When MAs are detected in many consecutive PPG segments, our system attributes them to near-wrist activities and then perform *MA Removal* to eliminate the impacted PPG segments. On the contrary, if MAs are detected in scattered or only a few consecutive segments, our system associates them with far-wrist activities and performs *MA Mitigation* to reconstruct related pulse waveforms. After the *Motion Artifact (MA) Filtering*, the data processing of our system is separated into two phases: *Training Phase* and *Authentication Phase*.

Training Phase. In this phase, our system performs *General Fiducial Feature Extraction* to extract the unique cardiac features from the PPG segment and its second derivative. This process applies to both wrist PPG measurements and fingertip ones. Next, we perform *Binary Gradient Boosting Classifier Construction* to train a binary classifier for each user. In particular, we construct a user's profile based on some extracted features and use the Gradient Boosting Tree (GBT) in training the classifier when the user enrolls in the system. Furthermore, our system regularly updates the classifier with new training data to accommodate PPG drifts over time in *Adaptive Updating*.

Authentication Phase. In the *Authentication Phase*, our system collects PPG segments in real-time and determines whether a current user is legitimate based on the PPG segments in a sliding window. Specifically, after our system filters MAs out from the PPG segments, it would further extract

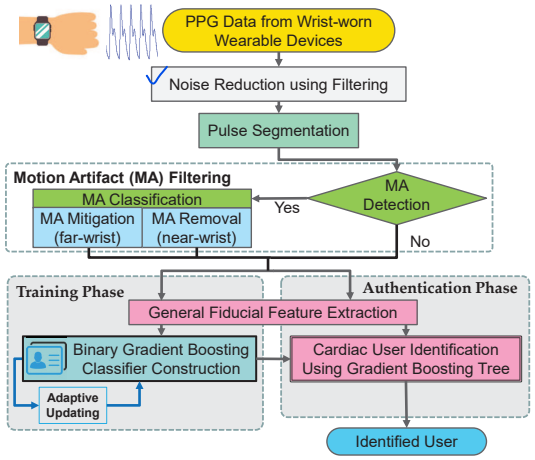


Fig. 5. Architecture of TrueHeart.

general fiducial features. Then our system performs *Cardiac User Identification Using Gradient Boosting Tree* by using the binary gradient boosting classifiers generated in the training phase to determine the user's identity based on each PPG segment. Finally, our system utilizes a majority-vote rule on the classified results of the PPG segments in the sliding-window to perform CA. In addition, our CA system is suitable for commodity wearable devices since their PPG sensors consume low power (e.g. 4mA) compared to battery capacities of these devices.

D. Challenges

Accurate Sensing Using Low-cost PPG Sensor on the Wrist. The low-cost PPG sensors in commodity wearable devices collect data from users' wrists at lower sampling rates with more noise and lower resolution. This will reduce the accuracy in user authentication.

Robust CA with Body Movements in Daily Activities. The PPG sensors in the wrist-worn wearable device are particularly susceptible to daily physical activities. Therefore we need to explore characteristics of MAs from the PPG measurement and develop technologies to effectively reduce such impacts.

Effective Feature Set for General PPG Measurements. The PPG measurements from the wrist area are unstable and weak, leading to fewer detectable fiducial features. Thus, we need to exam general effective features for CA.

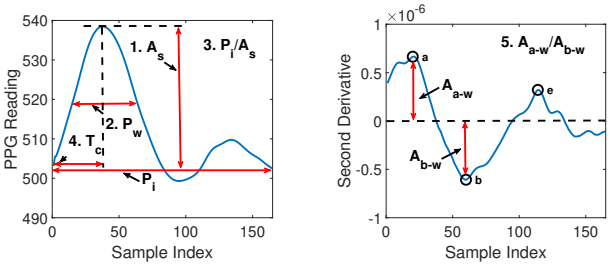
Persistent User Authentication Against PPG Drifts. The typical system-drifts in PPG sensors which could significantly impact the CA performance. Our system should study these drifts and adaptively accommodate the resulting PPG variations during a long time period.

IV. PPG Feature Extraction and User Authentication

In this section, we explore the cardiac features extracted from PPG measurements and present the details of our adaptive user authentication using gradient boosting.

A. General Wrist PPG Feature Extraction

We have shown that the PPG measurements from the wrist area have fewer fiducial features and non-fiducial features



(a) Raw PPG measurements

(b) Second derivative of raw PPG measurements

Fig. 6. Illustration of the five general fiducial features extracted from the raw PPG measurements and its second derivative.

compared to the PPG measurements from the fingertip. Therefore, we explore the fiducial features that are still available in the PPG measurements from the wrist area based on the 29 fiducial features that have been used for user authentication [16], [24].

General Wrist PPG Fiducial Features. Based on our experiments with 20 participants, we find that 60% of the PPG measurements from the wrist area have only one obvious systolic peak in a cardiac cycle. To let our CA system generally work for various types of PPG measurements, we select to use five fiducial features that only require a single systolic peak in the PPG measurements. Figure 6 illustrates how to derive the five fiducial features from the critical landmarks in the PPG pulse waveform. The five fiducial features are generally effective for the user authentication because they are always available regardless of the source of the PPG measurements (i.e., from the wrist area or the fingertip), and they have the physiological relationships with human cardiac systems. We summarize the five fiducial features and their physiological meanings as shown in Table I. Note that the five general fiducial features are always available in the PPG measurements from the fingertip. Therefore, our CA system is also applicable to the clinical PPG measurements. We provide a detailed evaluation of our system on both our PPG data from the wrist area and the fingertip in Section VII.

B. Adaptive cardiac authentication using Gradient Boosting Tree

Next, we build the binary classifier using Gradient Boost Tree (GBT) for user authentication. Comparing to other machine learning methods, GBT can handle the mixed types of the features with different scales, which is exactly what our general fiducial feature set possesses. Moreover, GBT is robust against the outliers via the robust loss functions and can eliminate the requirement of normalizing or whitening the feature data before classification [25].

Specifically, given N training samples $\{(x_i, y_i)\}$, where x_i and y_i represent the cardiac-related feature set and the corresponding identity label of the user (i.e., y_i = 1 or -1 represents whether x_i is from the current legitimate user), GBT seeks a function $\phi(x_i) = \sum_{m=1}^{M} \omega_m h_m(x_i)$ to iteratively select weak learners $h_m(\cdot)$ and their weights ω_j to minimize a loss function as follows:

$$\mathbf{L} = \sum_{i=1}^{N} L(y_i, \phi x_i). \quad (1)$$

TABLE I
LIST OF GENERAL WRIST PPG FEATURES.

Feature Name	Feature Description
Systolic Amplitude (A_s)	related to the stroke volume and directly proportional to vascular distensibility, which is distinguishable among different people.
Pulse Width (P_w)	the width of the PPG signal at the half-height of the systolic peak, and it correlates with the systemic vascular resistance.
Ratio of Pulse Interval to Systolic amplitude (P_i/A_s)	reflects the functionality of a person's cardiovascular system.
Crest Time (T_c)	indicates the pulse wave velocity, which is distinct from person to person.
Ratio of Amplitude of b-wave and a-wave (A_{b-w}/A_{a-w})	reflects the arterial stiffness and the distensibility of the peripheral artery, which are also different among people. In addition, this feature can also reflect the healthy level of different people.

We specifically adopt the GBT implementation from the SQBlib library [26] for cardiac-related feature training. In order to optimize the speed and accuracy of the GBT model, we empirically choose the exponential loss $L = e^{y_i \phi(x_i)}$ as the loss function $L(\cdot)$ with enough shrinkage (i.e., 0.1) and number of iterations (i.e., $M = 2000$), and we take a fraction of 0.5 as the sub-sampling of the training dataset. Once we have determined the loss function, next we will construct a binary gradient classifier $b_k(\cdots)$ for each user $g_k, k = 1, \cdots, K$ to complete the *Training Phase*. Then for the testing feature set, each binary gradient classifier will output a score. The reason to use binary classifier is that binary classifier has higher accuracy in differentiating one user from other users [27] which exactly meets the fundamental requirement of a CA system.

GBT에서 학습률 0.1, 트리 개수는 M개
트리마다 독립적으로 50% 서브샘플을 뽑습니다

In the *authentication phase*, our system utilizes the already built binary classifiers for all the users in parallel to classify incoming cardiac-related feature set x. In particular, we will obtain different confidence scores from each binary classifier, and choose the identity k of the binary classifier $b_k(x)$ with the highest score as the final classification. After the user classification, we adopt a non-overlapped sliding window-based approach to perform CA. In particular, we consider P continuous PPG segments in a sliding window as a basic CA unit and use the majority vote from the classification results of these PPG segments to determine the user's identity periodically. If equal or more than half of the PPG segments in the window are classified to be the same user, the system would allow the current user to pass the user authentication. Otherwise, the current user does not pass the user authentication. Unless mentioned elsewhere, we use the set the sliding window size to 4 PPG segments, which generally provides good performance as shown in our evaluation.

Adaptive Updating. We find that people's pulse patterns may slightly vary during the day. Therefore, we design our system to re-train the underlying classifier based on the recently collected PPG measurements after each successful user authentication. Specifically, our system regularly adds a small amount of the user's PPG measurements (e.g., 2min) to the training data to re-train a new classifier for the user in the background. This re-training process will stop until the new classifier meets the performance requirement (e.g., when the

CA accuracy reaches 90%), and the new classifier will take effect until the next time re-training process starts.

V. Motion Artifacts Detection and Filtering

In this section, we present the MA detection and classification methods. Based on different causes of MA, we present the details of the MA removal and MA mitigation.

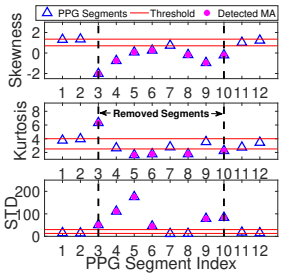
A. Motion Artifacts Detection

After the pulse segmentation mentioned in Section VI, the system first needs to detect whether MA is affecting the PPG segments or not. We find that when there is no MA, the PPG segments should contain similar pulse waveform, thus the statistics of each PPG segment should be stable over time. However, when the PPG segments are affected by MA, the statistics of PPG measurements vary a lot. Therefore, we propose to examine the statistics of each PPG segment and use a threshold-based approach to detect the existence of MA.

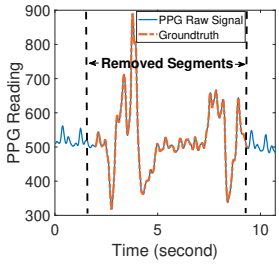
In particular, we choose three types of statistics (i.e., kurtosis, skewness, and standard deviation (STD)) efficiently measuring the symmetry, tails, and dispersion of the PPG segments respectively, which are used to effectively detect MA in existing work [28]. For each type of statistics, we derive its cumulative distribution function (CDF) based on high-quality PPG segments (about 20 seconds) without MA. From the CDF, we determine two thresholds that can include 95% of the values of particular statistics. The statistics of the testing PPG segments will be compared to the thresholds, respectively. If any of the three types of statistics from a PPG segment is out of the range determined by the corresponding two thresholds, the PPG segment is determined to be affected by MA. Figure 7 presents an example of our MA detection, which shows that our method can successfully detect the PPG segments affected by MA through the three types of statistics of the PPG segments in a sliding window. We note that the accuracy of our MA detection method is over 95% in our data from the wrist collected in the moving-scenario as described in Section VII.

B. Motion Artifacts Classification

The far-wrist activities (e.g., moving the forearm to reach a cup) usually create sparse and mild MA to PPG measurements, while the near-wrist activities (e.g., grabbing a cup) result in much stronger MA for a considerably longer period. Based on this observation, we develop an MA classification method, which examines the proportion of the PPG segments affected by MA in the sliding window $\mathcal{W}$ and determines whether the cause of MA is the near-wrist activities or far-wrist activities using a threshold-based approach. More specific, we denote the number of PPG segments that are determined to be affected and not affected by MA in the sliding window as M_W and N_W, respectively. The proportion of the PPG segments affected by MA in the sliding window is defined as the ratio $\lambda = \frac{M_W}{N_W}$. Next, λ is compared to a threshold θ_{ma}. The cause of MA is classified as the near-wrist activities if the $\lambda \geq \theta_{\text{ma}}$. Otherwise, the cause of MA is classified as the far-wrist activities. From our experimental results from all 20 participants, we find that a short time period $\mathcal{W} = 10$s is sufficient to cover the duration of typical arm movements, and the threshold $\theta_{\text{ma}} = 30\%$ is general enough to provide high accuracy of categorizing the arm movements for all participants. In our evaluation, we apply this general threshold for categorizing the movements.



(a) MA detection in a sliding window for the near-wrist activity



(b) The PPG measurements after MA detection and removal for the near-wrist activity

Fig. 7. Performance of MA detection and MA removal for the near-wrist activity.

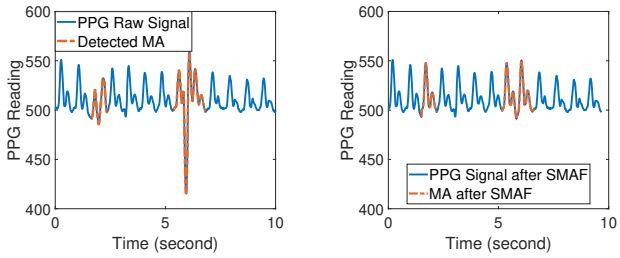
C. Motion Artifacts Removal for Near-wrist Activities

When the system determines that the PPG segments are affected by the near-wrist activities, it implies that the PPG measurements are significantly distorted by the MA during the time in the sliding window, which we consider them unrecoverable. In this case, we remove all the PPG segments affected by MA and only perform user authentication using the rest of the PPG segments in the sliding window. However, we find that the PPG segments affected by MA may not be continuous, and the interval between two affected segments may be too short (e.g., $1 \sim 2$ seconds including $1 \sim 3$ PPG segments) for extracting a complete pulse waveform that can be used to perform user authentication. Hence, we remove all the PPG segments in between the first and last segments affected by MA and keep the unaffected PPG segments for user authentication.

An example of our MA removal for the near-wrist activity is shown in Figure 7 (a). Based on the MA detection results (i.e., 7 out of all the 12 PPG segments are determined as MA), we can determine the PPG measurements in the sliding window contains the near-wrist activity. Thus, our system removes the PPG segments affected by MA (i.e., from PPG segment index $3 \sim 10$) between the first and last detected MA in this sliding window. As shown in Figure 7 (b), our MA removal method can successfully remove the PPG measurements that are impacted by the near-wrist activities with respect to the ground truth. In addition, it should be noted that our CA system could still authenticate the user when the hand is stable before/after the near-wrist activities, and removing the MA caused by the near-wrist activities does not influence the user experience since user authentication can be done before the near-wrist activities.

D. Motion Artifacts Mitigation for Far-wrist Activities

When the system determines that the PPG segments are affected by the far-wrist activities, we notice that the interference of MA is usually small and recoverable. Therefore, we employ



(a)Raw PPG measurements and detected MA

(b)PPG measurements after SMAF filtering

Fig. 8. Example of the MA mitigation using SMAF.

a special moving average filter (SMAF) to mitigate those MA and retain them for continuous authentication. The basic idea is to average each recognized MA with several pure pulse segments (i.e., the typical PPG segments without MA) of the current testing user. Then the MA is able to be mitigated from the averaged results. Specifically, we first align the pure pulse PPG segments using the systolic peaks in order to maintain the locations of the critical fiducial points. Since the number of the samples in each pulse segment is not equal, we then interpolate those PPG segments to make them have the same length. After the interpolation, we will apply the SMAF on the pure pulse segments and MA using the following equation:

$$\mathbf{S} = \frac{\sum_{h=1}^{N} \overrightarrow{P_h} + \overrightarrow{M}}{N+1}, \tag{2}$$

where the $\overrightarrow{P_h}$ represents the pure pulse segments, $\overrightarrow{M}$ is MA that requires the mitigation, and totally N pure pulse segments and 1 MA are averaged with the mitigated result as $\mathbf{S}$. In particular, we use 4 pure pulse segments for the proposed SMAF. After the SMAF, we use the smooth function to ensure the continuity of the filtered signal. Figure 8 illustrates the effectiveness of our MA mitigation for a far-wrist activity, which is raising forearm to check the time. From Figure 8 (a), we can see that a small proportion of PPG measurements (i.e., the measurements highlighted in red) are affected by the far-wrist activities and detected by our MA detection method. Figure 8 (b) presents the results after applying the SMFA filter on the PPG measurements, which can mitigate the impact and reconstruct the pulse waveforms.

VI. PPG Data Preprocessing and Segmentation

A. Data Preprocess

The PPG measurements from the low-cost PPG sensor in wrist-worn wearable devices inevitably contain baseline drift and high-frequency interference. Since the frequency of the pulsatile component in PPG is $0.5-4$Hz, and the frequency of MA is 0.1Hz and above, our system firstly applies a band-pass filter to reduce the effect of the baseline drift and high-frequency noise. In particular, we implement a Butterworth bandpass filter with the passband $0.5-6$Hz and the order as 2 to only retain the pulsatile components together with the MA components having a similar spectrum.

B. Pulse Segmentation

Our system determines the starting and ending points of all the PPG segments in the sliding window. Figure 2 shows that the starting and ending points of a typical complete cardiac cycle correspond to the two valley points before the systolic and after diastolic points, respectively. Ideally, we can find all the valley points in the sliding window and extract the data between every two valley points as the PPG segments. However, we find that the dicrotic notch could have the lowest amplitude (i.e., "fake" valley) in the cardiac cycle. Particularly, we tackle this issue based on the fact that the time distances from the systolic peak to the starting and ending points are in the range of $T_s = 0.15$s~ 0.26s and $T_e = 0.44$s~ 0.74s, respectively [29]. Therefore, the accurate PPG segment can be extracted by selecting the valleys that are within the typical time ranges T_s and T_e before and after each systolic peak, respectively. In addition, through our experiments with 20 participants, we empirically determine the sliding window as 2s larger than one typical pulse waveform (e.g., $0.6 \sim 1$ second) to ensure the effectiveness and accuracy of the PPG segmentation. We also note that our segmentation method is effective with MA because the system finds PPG segments in the sliding window based on the peaks and valleys that fulfill the criteria even though the waveform may be distorted.

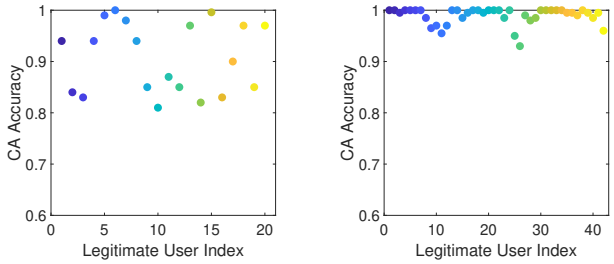
Fig. 9. Prototype: wrist-worn PPG sensing platform.

VII. Performance Evaluation

A. Experimental Methodology

Wearable Prototype. We notice that existing commodity wearable devices can only provide the computed heart rate instead of direct access to raw PPG readings. Therefore, we design a wrist-worn PPG sensing prototype as shown in Figure 9, which refers to the layout of PPG and motion sensors in commodity wrist-worn wearable device (e.g., Apple Watch). Specifically, the prototype consists of one commodity green LED PPG sensor attached to the inner side of the wristband and a motion sensor (i.e., accelerometer) attached to the outside of the wristband. These sensors are connected to an Arduino UNO (REV3) board for the sensor measurements acquisition, which is under a $300Hz$ sampling rate. The PPG measurements are transferred to a laptop (i.e., Dell Latitude E6430) to perform user authentication.

Data Collection. We recruit 20 healthy participants whose ages are between 20 to 40 to collect PPG measurements using our wearable prototype. Two different scenarios are adopted to evaluate our system for various practical application scenarios: In the *static scenario*, 20 participants are asked to sit quietly for 10 mins, respectively. While in the *moving scenario*, we ask 5 participants to perform the *far-wrist activities* (i.e., moving the forearms) and the *near-wrist activities* (i.e., grabbing up a cup and drinking water) repeatedly for 2 mins and sit still for 3 mins. In total, we collect around $15,000$ PPG pulse segments from the wrist in the static-scenario and $4,200$ pulse



(a) PPG from the wrist area (b) PPG from the fingertip

Fig. 10. CA accuracy of TrueHeart using the PPG measurements from the wrist areas and the fingertips.

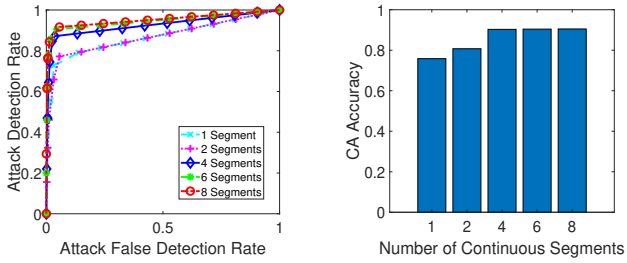


Fig. 11. ROC curves under the random attack.

Fig. 12. Performance with different lengths of the sliding window.

segments in the moving-scenario, respectively. In addition, we also test our system on the IEEE TBME Benchmark dataset [30], which has 8-mins PPG data collected from the fingertips of 42 people with a sampling rate of 300 Hz.

B. Evaluation Metrics

Our system periodically authenticate the user based on the PPG segments in a sliding window and labels the sliding window as the user or attacker, respectively. We define our evaluation metrics as follows:

CA Accuracy. The number of sliding windows that are correctly labeled as the user over the total number of sliding windows that are examined during the CA process.

Attack Detection Rate. The number of sliding windows that are correctly labeled as the attacker over the number of sliding windows that are associated with the attacker during the CA process.

Attack False Detection Rate. The number of sliding windows that are incorrectly identified as the attacker over the number of sliding windows that are associated with the user during the CA process.

Receiver Operating Characteristic (ROC) Curve. It reflects the trade-off between Attack Detection Rate and Attack False Detection Rate. The smallest distance from the point on the ROC curve to the top-left corner corresponds to the optimum model.

In our evaluation, 20 rounds of Monte Carlo cross-validation are employed for the 10-mins of the collected user data, among which 5-mins for training and the rest of the data for authentication.

C. Continuous Authentication (CA) Performance

We first evaluate the general performance of TrueHeart by examining the CA accuracy in the static scenario. In particular, we consider each participant acts as a legitimate user once while remaining participants act as attackers. Figure 10(a) shows that each user achieves comparable high CA accuracy with an average of 90.73% CA accuracy, which indicates that TrueHeart can successfully authenticate users with high accuracy using the wrist-worn wearable devices. In addition, Figure 10(b) shows that our system can achieve even better performance on the PPG data from the fingertip [30] with 39 out of 42 people having the CA accuracy above 96%. This is because the PPG measurements from the fingertip are stronger and stabler than the wrist area. These results not only demonstrate the promising practical usability of our proposed user authentication system on common wrist-worn wearable devices but also indicate that it has promising usage in clinical environments such as telemedicine and smart-health applications.

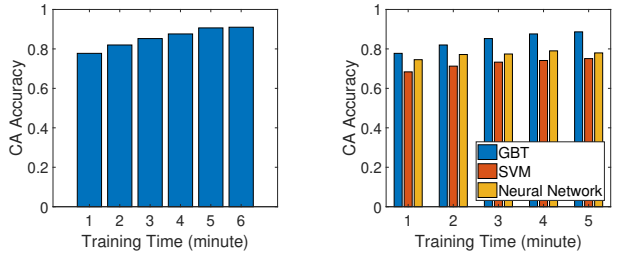
Moreover, to study the performance of our system when defending against the random attack, Figure 11 shows that the ROC curve gets closer to the point $(0, 1)$ when the number of the PPG segments in a sliding window becomes larger. Particularly, our attack detection rate reaches to over 88% with the attack false detection rate of around 3.9% when the length of the sliding window is 4. And our system can achieve over 90% attack detection rate and less than 4.2% attack false detection rate with six or more PPG segments in a sliding window. Those results show that our CA system is robust against the random attacks.

D. Impact of Various Factors

Impact of the Sliding Window Length. The length of the sliding window corresponds to the number of continuous PPG segments to perform the majority vote for user authentication. Particularly, we test the different lengths of the sliding window with 1, 2, 4, 6, 8 continuous PPG segments (i.e., about 0.7s, 1.4s, 3s, 4.4s, and 6s). Figure 12 shows the CA accuracy increases as the increment of the sliding window length and becomes stable at about 90% with four or more PPG segments. Therefore, we adopt the sliding window with 4 continuous PPG segments in our system, which not only provides the high CA accuracy but also has the short response time for the authentication (i.e., around 3s).

Impact of Training Data Size. Since the training data size influences the ease of use in terms of the time for data collection, so we particularly test 1, 2, 3, 4, 5, and 6 mins' static PPG signals of each user for training respectively, and use the rest data for testing. Figure 13(a) shows that an average CA accuracy of 77.75% is achieved only using 1 min's data of each user for training. Moreover, the average CA accuracy can increase to 90.65% and becomes stable when using 5-mins or more training data of each user. Those results prove that our system is suitable for practical use since it can achieve very high CA accuracy with the only limited size of training data (e.g., 5-mins per user).

Impact of Machine Learning Methods. We study the performance of our system with different underlying machine learning models. Specifically, we adopt the support vector machine (SVM) and neural network (NN) using the LIBSVM



(a) Performance with different sizes of the training data

(b) Performance with different machine learning methods

Fig. 13. The impacts of the training size and the machine learning method.

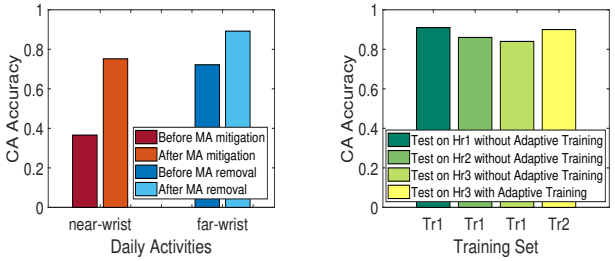


Fig. 14. Performance of MA removal.

Fig. 15. Performance comparison with different testing data with and without adaptive training.

library [31] and the multi-layer perceptron in Scikit-learn [32], respectively. Figure 13(b) shows that GBT has the best CA accuracy of 90% compared with SVM (scaling the data) and NN whose CA accuracy is 75% and 80% respectively. This result indicates that GBT easily tuned with flexible optimization options is more suitable for our CA system than the machine learning methods which either are difficult to determine the appropriate kernel (e.g., SVM) or require a large amount of training data and expertise to tune the model (e.g., NN).

Impact of Sampling Rate. The sampling rate affects the power consumption and computational cost in the wearable devices. In particular, we find that the CA accuracy is as high as 88% at the lowest sampling rate (i.e., $25Hz$) and increases slightly with the increased sampling rate and becomes stable with 90.7% CA accuracy since $100Hz$. Those findings imply that our CA system is not only compatible with the commodity wrist-worn wearable devices (e.g., Samsung Simband [33] adopts $128Hz$ PPG sampling rate) but also supports the hardware with even lower PPG sampling rate.

E. CA Performance with MA Removal and MA Mitigation

We next study the performance of our MA removal method on near-wrist activities and MA mitigation method on far-wrist activities among 5 participants, respectively. As shown in Figure 14, while performing far-wrist activities such as moving forearm, our system could still achieve 72.2% CA accuracy even without applying the MA mitigation method and the CA accuracy increases to 89.2% after MA mitigation. Furthermore, we can see that our system has the CA accuracy as 36.6% before MA removal and achieve 75.2% after MA removal for the near-wrist activities such as grabbing up a cup to mimic drinking water gesture. Those results show that the far-wrist activities have a relatively slight impact on our CA system, whereas the near-wrist activities have more impacts due to the involvement of the tendon and muscle in the wrist area. Overall, our system has a decent performance after applying the MA removal method on the near-wrist activities and MA mitigation method on the far-wrist activities, which implies that it's practical for daily life usage.

F. Effectiveness of Adaptive Training

We evaluate our adaptive training using the data collected by one user across three different hours in a day. Specifically, we collect 1-hour PPG data starting at 11 AM, 1 PM, and 4 PM, respectively. In Figure 15, $Tr1$ represents the training set is only from the first hour and Tr_2 represents the mixed training set includes the data from both the first hour and 2 mins' data from the third hour. We can see that our system trained by $Tr1$ can achieve 91% CA accuracy during the first hour, and decreases 5% during the second hour and 7% during the third hour, respectively. These results demonstrate that the user cardiac system indeed has some fluctuations during a long-time period that slightly impact the CA performance. Moreover, after the adaptive retraining with Tr_2, the CA accuracy will increase back to 90% during the third hour. Those results prove that our system is suitable for long-time user authentication with few times of adaptively retraining which requires a very small amount of the new data. (e.g., routinely retrain every 3 hours with only 2 mins' new data).

VIII. Conclusion

In this paper, we develop a low-cost PPG-based continuous user authentication (CA) system, TrueHeart, using the wrist-worn wearable devices. Specifically, we explore the diverse PPG measurements among 20 participants and determine the representative and general fiducial feature sets that can facilitate our CA system. We develop an effective motion artifact (MA) detection method based on the statistics of the PPG segments. In addition, MA classification and MA removal modules are designed to mitigate the impact of body movements. To ensure the long-term robustness of our CA system, we develop an adaptive user authentication method using the gradient boosting tree (GBT) technique. We devise a wrist-worn PPG sensing prototype and conduct extensive experiments with 20 participants under static and different moving scenarios. The results show that our system can achieve a high average CA accuracy of over 90% and a low attack false detection rate of 4% in practice. We are aware that continuous near-wrist activity and unexpected sickness would cause drastically cardiac status changes and impact the performance of our system. In those cases, our system would notify the users of using the tradition authentication approach (e.g., password) to verify their identity temporarily, then update itself using the adaptive learning.

Acknowledgment

This work was partially supported by the National Science Foundation Grants CNS1566455, CNS1826647, CNS1954959, CCF1909963, CCF2000480, and ARO Grant W911NF-18-1-0221.

REFERENCES

[1] A. H. Lashkari, S. Farmand, D. Zakaria, O. Bin, D. Saleh *et al.*, "Shoulder surfing attack in graphical password authentication," *arXiv preprint arXiv:0912.0951*, 2009.

[2] A. J. Aviv, K. L. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens." *Woot*, vol. 10, pp. 1–7, 2010.

[3] A. Bhargav-Spantzel, A. C. Squicciarini, S. Modi, M. Young, E. Bertino, and S. J. Elliott, "Privacy preserving multi-factor authentication with biometrics," *Journal of Computer Security*, vol. 15, no. 5, pp. 529–560, 2007.

[4] A. Ometov, S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen, and Y. Koucheryavy, "Multi-factor authentication: A survey," *Cryptography*, vol. 2, no. 1, p. 1, 2018.

[5] A. Al Abdulwahid, N. Clarke, I. Stengel, S. Furnell, and C. Reich, "A survey of continuous and transparent multibiometric authentication systems," in *European Conf. on Cyber Warfare and Security*, 2015, pp. 1–10.

[6] I. Traore, I. Woungang, M. S. Obaidat, Y. Nakkabi, and I. Lai, "Combining mouse and keystroke dynamics biometrics for risk-based authentication in web environments," in *Proceedings of the Fourth International Conference on Digital Home Digital Home (IEEE ICDH)*, 2012, pp. 138–145.

[7] S. Mare, A. M. Markham, C. Cornelius, R. Peterson, and D. Kotz, "Zebra: Zero-effort bilateral recurring authentication," in *Security and Privacy (SP), 2014 IEEE Symposium on*. IEEE, 2014, pp. 705–720.

[8] Y. Ren, Y. Chen, M. C. Chuah, and J. Yang, "User verification leveraging gait recognition for smartphone enabled mobile healthcare systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 9, pp. 1961–1974, 2015.

[9] S. J. Kang, S. Y. Lee, H. I. Cho, and H. Park, "Ecg authentication system design based on signal analysis in mobile and wearable devices," *IEEE Signal Processing Letters*, vol. 23, no. 6, pp. 805–808, 2016.

[10] J. R. Pinto, J. S. Cardoso, A. Lourenço, and C. Carreiras, "Towards a continuous biometric system based on ecg signals acquired on the steering wheel," *Sensors*, vol. 17, no. 10, p. 2228, 2017.

[11] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren, "Cardiac scan: A non-contact and continuous heart-based user authentication system," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (ACM MobiCom)*, 2017, pp. 315–328.

[12] A. Bonissi, R. D. Labati, L. Perico, R. Sassi, F. Scotti, and L. Sparagino, "A preliminary study on continuous authentication methods for photoplethysmographic biometrics," in *Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2013 IEEE Workshop on*. IEEE, 2013, pp. 28–33.

[13] A. Sarkar, A. L. Abbott, and Z. Doerzaph, "Biometric authentication using photoplethysmography signals," in *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*. IEEE, 2016, pp. 1–7.

[14] N. Karimian, M. Tehranipoor, and D. Forte, "Non-fiducial ppg-based authentication for healthcare application," in *Proceedings of the 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)(IEEE EMBS)*, 2017, pp. 429–432.

[15] N. Karimian, Z. Guo, M. Tehranipoor, and D. Forte, "Human recognition from photoplethysmography (ppg) based on non-fiducial features," in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)*, 2017, pp. 4636–4640.

[16] A. R. Kavsaoğlu, K. Polat, and M. R. Bozkurt, "A novel feature ranking algorithm for biometric recognition with ppg signals," *Computers in biology and medicine*, vol. 49, pp. 1–14, 2014.

[17] M. A. S. Mondol, I. A. Emi, S. M. Preum, and J. A. Stankovic, "User authentication using wrist mounted inertial sensors," in *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2017, pp. 309–310.

[18] P. Casale, O. Pujol, and P. Radeva, "Personalization and user verification in wearable systems using biometric walking patterns," *Personal and Ubiquitous Computing*, vol. 16, no. 5, pp. 563–580, 2012.

[19] A. Rahman, V. Lubecke, O. Boric-Lubecke, J. Prins, and T. Sakamoto, "Doppler radar techniques for accurate respiration charac-terization and subject identification," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2018.

[20] M. Guennoun, N. Abbad, J. Talom, S. M. M. Rahman, and K. El-Khatib, "Continuous authentication by electrocardiogram data," in *Proceedings of the 2009 IEEE Toronto International Conference Science and Technology for Humanity (IEEE TIC-STH)*, 2009, pp. 40–42.

[21] C. Camara, P. Peris-Lopez, L. Gonzalez-Manzano, and J. Tapiador, "Real-time electrocardiogram streams for continuous authentication," *Applied Soft Computing*, vol. 68, pp. 784–794, 2018.

[22] X. Niu, H. Han, S. Shan, and X. Chen, "Continuous heart rate measurement from face: a robust rppg approach with distribution learning," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 642–650.

[23] J. T. Shepherd and P. M. Vanhoutte, "The human cardiovascular system. facts and concepts," 1979.

[24] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current cardiology reviews*, vol. 8, no. 1, pp. 14–25, 2012.

[25] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning new york," *NY: Springer*, 2009.

[26] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 526–533.

[27] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognition*, vol. 44, no. 8, pp. 1761–1776, 2011.

[28] L. Pu, P. J. Chacon, H.-C. Wu, and J.-W. Choi, "Novel tailoring algorithm for abrupt motion artifact removal in photoplethysmogram signals," *Biomedical Engineering Letters*, vol. 7, no. 4, pp. 299–304, 2017.

[29] T. Bombardini, V. Gemignani, E. Bianchini, L. Venneri, C. Petersen, E. Pasanisi, L. Pratali, D. Alonso-Rodriguez, M. Pianelli, F. Faita *et al.*, "Diastolic time–frequency relation in the stress echo lab: filling timing and flow at different heart rates," *Cardiovascular ultrasound*, vol. 6, no. 1, p. 15, 2008.

[30] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–1953, 2013.

[31] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[33] Simband. (2017) Why is 128 hz used as a sampling frequency for the ppg signals? [Online]. Available: http://www.simband.io/documentation/faq.html