Taming Irregular Cardiac Signals for Biometric Identification

Wang, Weizheng; Wang, Qing; Zuniga, Marco

DOI
10.1145/3624570

Publication date
2023

Document Version
Final published version

Published in
ACM Transactions on Sensor Networks

Taming Irregular Cardiac Signals for Biometric Identification

WEIZHENG WANG, QING WANG, and MARCO ZUNIGA, Delft University of Technology, The Netherlands

Cardiac patterns are being used to provide hard-to-forge biometric signatures in identification applications. However, this performance is obtained under *controlled scenarios* where cardiac signals maintain a relatively uniform pattern, facilitating the identification process. In this work, we analyze cardiac signals collected in more *realistic (uncontrolled) scenarios* and show that their high signal variability makes them harder to obtain stable and distinct features. When faced with these irregular signals, the state-of-the-art (SOTA) reduces its performance significantly. To solve these problems, we propose the CardioID framework[1] with two novel properties. First, we design an adaptive method that achieves stable and distinct features by tailoring the filtering process according to each user's heart rate. Second, we show that users can have multiple cardiac morphologies, offering us a bigger pool of cardiac signals compared to the SOTA. Considering three *uncontrolled* datasets, our evaluation shows two main insights. First, while using a PPG sensor with healthy individuals, the SOTA's balanced accuracy (BAC) reduces from 90–95% to 75–80%, while our method maintains a BAC above 90%. Second, under more challenging conditions (using smartphone cameras or monitoring unhealthy individuals), the SOTA's BAC reduces to values between 65–75%, and our method increases the BAC to values between 75–85%.

CCS Concepts: • **Security and privacy → Biometrics**; • **Human-centered computing → Smartphones;**

Additional Key Words and Phrases: Irregular cardiac signals, biometric identification, authentication, uncontrolled scenarios

ACM Reference format:
Weizheng Wang, Qing Wang, and Marco Zuniga. 2023. Taming Irregular Cardiac Signals for Biometric Identification. *ACM Trans. Sensor Netw.* 20, 1, Article 25 (December 2023), 26 pages.
https://doi.org/10.1145/3624570

1 INTRODUCTION

Biometrics plays a fundamental role in human identification, and the most popular systems rely on external features, such as fingerprints, iris patterns, and face contours. These systems have excellent precision but they are vulnerable to attacks: fingerprints can be recreated in latex from

[1]This work was partially presented at the International Conference on Embedded Wireless Systems and Networks (EWSN 2022) [44].

Authors' address: W. Wang, Q. Wang, and M. Zuniga, Delft University of Technology, Van Mourik Broekmanweg 6, Deflt 2628 XE, The Netherlands; e-mails: {w.wang-14, qing.wang, m.a.zunigazamalloa}@tudelft.nl.

1550-4859/2023/12-ART25
https://doi.org/10.1145/3624570

touched objects [24]; iris patterns can be scanned and emulated [17]; and pictures from the Internet can be used to obtain renditions that can fool face recognition systems [4].

To overcome the fundamental weakness of *external* features, i.e., the fact that they can be easily captured because they are constantly exposed, researchers are investigating *internal* biometric signals, which are hidden under our skin, and hence, they are hard to obtain and forge.[2] An approach that is gaining interest is the use of cardiac patterns since they are uniquely defined by the heart, lung, and vein structures of an individual. These cardiac patterns can be obtained with a **photoplethysmogram** (**PPG**), which measures changes in blood volume via light absorption. PPG signals can be acquired with simple inexpensive sensors that are widely available on wearable devices. For example, one option is to use a pulse oximeter on a finger, which consists of a small LED and a simple photosensor [42]; another option is to place a finger on top of the flashlight and camera in a smartphone [28]. With both types of sensors, researchers have shown that PPG signals can provide between 85% and 95% identification accuracy for groups consisting of tens of people [15, 24, 28, 36].

Challenge. The results obtained so far for PPG identification are promising, but they have been obtained mainly under ideal situations: accurate sensors used in controlled environments. These two factors (sensors and environment) determine how similar cardiac cycles are for the same individual. The higher the similarity of the cardiac cycles, the higher the identification accuracy. We show that when PPG signals are gathered in a more natural (uncontrolled) manner, the cardiac cycles can be highly irregular, significantly decreasing the accuracy of **state-of-the-art** (**SOTA**) approaches.

Our contributions. Considering the above challenge, we analyze the pernicious effects of irregular cardiac cycles on biometric identification and propose a novel framework to overcome those effects. In particular, our work provides four main contributions:

Contribution 1: Morphology Stabilization [Section 4]. The biometric information present in cardiac cycles is restricted to a narrow spectrum of the signal. A key limitation of the SOTA approaches is that their filters target the *same* spectrum for *all* individuals. This one-size-fits-all approach leads to either information loss (if the default spectrum is too narrow for a particular individual) or insufficient noise filtering (if the spectrum is too broad). We propose an adaptive technique that fine-tunes the filtering parameters based on the individual cardiac properties. This approach allows us to obtain more stable and distinctive features per user.

Contribution 2: Morphology Classification [Section 5]. SOTA studies assume that the cardiac pulses of individuals have a *single* dominant morphology (shape). Assuming a single morphology means that several "non-conforming" cardiac periods can be unnecessarily discarded, affecting the responsiveness of the system. More importantly, we find out that in some cases, the strict SOTA assumption of considering a single dominant morphology, leaves out users that rarely have such cardiac morphology, rendering the SOTA methods futile for those users. We show that a *single user* can have *multiple* valid morphologies. Our ability to consider a wider range of morphologies reduces the system's response time, increases user inclusion (to serve more people), and facilitates identifying the rightful individual even when his/her cardiac periods are different from each other.

Contribution 3: Analysis of non-linear effects [Section 6]. The SOTA utilizes PPG signals to perform two types of biometric applications: identification and authentication. For identification, the SOTA uses linear (PCA [28, 48], LDA [36, 46]) and non-linear approaches (NN-based [40, 42]), but there is no analysis determining what approach is better and why. We show that if we tackle the non-

[2] A few studies show that it is possible to spoof **electrocardiography** (**ECG**) authentication systems [12, 21], but this assumes access to a compromised medical database. That process is harder than obtaining a victim's fingerprint through the objects he/she has touched or obtaining a victim's face image by searching online.

Table 1. The Most Relevant Studies in the SOTA

	Application	Sensor Type	# Subjects	# Features	Decision Method	# PT	Accuracy
[15], 2003	Identification	pulse oximeter	17 (own)	4 (fiducial)	discriminant function	–	94,0%
[5], 2013	Ident.	pulse oximeter	44 14	1 (entire period)	correlation functions	16	5.3% (EER) 14.5% (EER)
[24], 2014	Ident.	pulse oximeter	30	40 (fiducial)	KNN	1	87.0% (F1)
[36], 2016	Ident.	pulse oximeter	32 , [25]	15 (fiducial)	LDA & QDA	5	92.5% (Rank-1)
[42], 2017	Identification	pulse oximeter	42 , [23]	DWT (non-fiducial)	SVM & NN	–	100%
[46], 2018	Auth.	pulse oximeter	42 , [23] 32 , [25] 41 , [1]	CWT (non-fiducial)	Direct-LDA	8-40	1.6% (EER) 3.8% (EER) 8.8% (EER)
[28], 2019	Auth.	camera	25	30 (fiducial) & 36 (non-fiducial)	PCA	1	95.8% (BAC)
CardioID	Both	Both	35 , [38] 43 10	32,38,44 (fiducial)	LDA & NN (Ident.) PCA (Auth.)	1	93.2% , 95.5% 78.2% , 85.3% 83.6% , 84.3% (BAC: Iden. , Auth.)

Some studies evaluate multiple datasets, if the dataset is public, a reference is given in the '# subjects' column. The studies in bold are used as comparison baselines in our work.

linear effects of cardiac cycles at an early stage, both approaches, linear and non-linear, render similar results. For authentication, we identify two main shortcomings in SOTA methods: the use of Euclidean distances and the assumption that the features of a subject form a *single* cluster. To ameliorate these non-linear effects, we propose a *multi*-cluster approach, together with the use of the Mahalanobis distance.

Contribution 4: Thorough multi-sensor and multi-application evaluations [Section 7]. The evaluation of cardiac signals for biometric applications can be divided into four quadrants: based on the type of sensor (pulse oximeter or camera) and application (identification or authentication). Most studies evaluate a single quadrant with healthy subjects (usually, identification with pulse oximeters), *no study has evaluated all four quadrants or considered unhealthy individuals.* Our evaluation assesses both applications relying on three datasets. The datasets consider both types of sensors (pulse oximeter and camera) and two types of individuals (healthy and unhealthy). Overall, the results show that in uncontrolled scenarios the average **balanced accuracy** (**BAC**) of the SOTA drops beyond 15%, depending on the complexity of the uncontrolled scenario. This is a significant drop for identification and authentication applications. For the less dynamic dataset (using pulse oximeter sensors with healthy individuals), our method improves the accuracy by 15%. For the most challenging datasets (using smartphone cameras or monitoring patients in ICU), our methods improve the BAC with values above 10%.

2 RELATED WORK

We divide the related work into three main phases, highlighting the elements we build upon from the SOTA and the novelty of our work. A summary of the most relevant studies is presented in Table 1.

Phase 1: Basic identification. Gu et al. report the first results for PPG identification using just four features. They achieve an accuracy of 94% using a discriminant function [15]. Later, researchers found that the derivatives of a PPG signal can provide more stable and unique features [47]. Motivated by those initial results, researchers performed further experiments and found that the reported high accuracy is strongly dependent on the data-gathering process. Spachos et al. [40] consider fiducial features with two datasets. They report widely different performances for each set, **Equal Error Rates** (**EERs**) of 0.5% vs. 25%, leading them to state that PPG signals can be used for identification "***given that [they] are collected under controlled environments and***

with accurate sensors". Bonissi et al. [5] also find significant differences in EER depending on the databases they use, 5.3% vs. 14.5%.

One of the most comprehensive evaluations is performed by Kavsaoğlu et al. [24]. They use 40 features from PPG signals and their derivatives (first and second), and utilize a *single period* for testing to obtain an F1 accuracy of 87.2% (F1 is a stricter metric compared to BAC). From that work, we borrow the idea of using the second derivative and multiple features. We implemented this method and used it as a comparison baseline for identification.

Phase 2: non-fiducial features and more challenging PPG signals. An important motivation for our study comes from [36] and [46]. Sarkar et al. [36] analyze PPG signals with subjects that undergo various emotions. To enforce the same morphology, they normalize signals in time and amplitude so fiducial features can maintain a common pattern. This approach, however, requires 20 cardiac periods per testing sample (~15 sec, too long of a delay). Yadav et al. [46] also look into PPG signals that consider different levels of emotions and physical exercise, but they propose to use *non*-fiducial features based on **continuous wavelet transform** (**CWT**). CWT considers the spectral response of a signal, which is more resilient to noise than geometric (fiducial) features, but they still require long testing sequences, between 8 and 40 cardiac periods (~6 to 60 sec). Karimian et al. [42] also utilize non-fiducial features (discrete wavelet transformation, DWT), and report an accuracy of 100%, compared to a 95–98% accuracy obtained with fiducial features, but does not report the number of periods used for testing.

From the above studies, we take two insights. First, the normalization of PPG signals in time and amplitude to overcome distortions caused by emotions. Second, we do not use spectral features due to the many cardiac periods needed for testing, but we do perform a thorough spectral analysis (harmonic filtering) to obtain more stable and distinct fiducial features.

Phase 3: cameras and authentication. Most studies focus on performing *identification* with *pulse oximeters*, but a recent work uses smartphone *cameras* to attain *authentication* [28] (CardioCam). Authentication is more challenging than identification because it trains the system with a single user. CardioCam achieves a high BAC (95.8%) using a single cardiac period for testing. Motivated by CardioCam, we also collect data with a smartphone camera and implement the *signal processing chain* proposed in that work (filters, features, and PCA method). Our results show that its performance decreases with irregular PPG signals. Furthermore, for some users, the requirements of CardioCam's morphology are so strict that they would not be able to use the system.

There are also *some other studies related to our work.*

Cardiac health applications. Several cardiac health applications use a smartphone camera. Chandrasekaran et al. [8] combine sound information from the chest and video from a fingertip to measure people's blood pressure. Their estimation accuracy is above 95%. HemaApp [43] infers hemoglobin levels based on the light absorption detected by a smartphone camera and achieves sensitivity and precision of 85.7% and 76.5%. Despite the differences in the application, we share with these studies the need to fine-tune cameras to obtain cardiac signals.

Identification with ECG sensors. The most common cardiac sensor is ECG. These sensors are widely available in hospitals and have also been used for identification. Safie et al. use ECG features to obtain an AUR and EER of 0.9101 and 0.1813, respectively [35]. Silva et al. explore a less invasive form of ECG sensors for authentication, finger-ECG [11]. They utilize pre-processed templates as inputs for K-NN and SVM and obtain an EER below 9.1%. Similarly, Singh et al. [39] use just one electrode to extract 19 fiducial features, and apply an adaptive threshold to perform authentication with a 99% accuracy. ECG sensors are more accurate than PPG sensors and smartphone cameras, but they are less pervasive and their filtering and identification methods are similar to the SOTA studies using pulse oximeters.

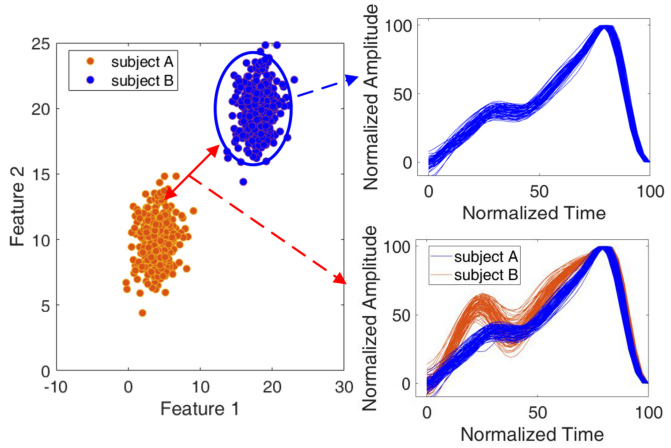


Fig. 1. Sample application with *controlled* PPG signals.

3 PRELIMINARIES

3.1 PPG Basics

The cardiac cycle represents the change in blood pressure determined by our hearts and blood circulation systems. Given that people have different heart structures in terms of volume, surface shape, and motion dynamics [7, 16, 27], and different tissue thickness and blood vessel distribution [16], the cardiac signal has been used to obtain unique biometric signatures [19, 48]. A cardiac cycle can be measured in various manners, the simplest option is to obtain a PPG signal by measuring the amount of light absorbed by our body as blood flows through. A PPG signal can be measured with sensors containing inexpensive LEDs and photodiodes, or with the flashlight and camera in smartphones. The geometric relations among the various peaks and valleys present in a PPG signal (heights, widths, etc.) [24], or the spectral information in the frequency domain [42], are optional features used to perform identification.

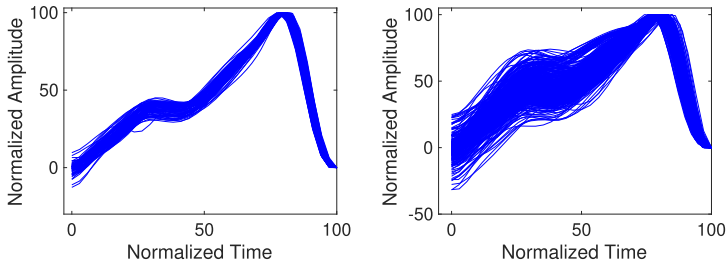
3.2 Applications, Morphologies, and Metrics

We analyze the performance of two applications: identification and authentication. In ***identification***, the population size is *known* and the training phase requires gathering data from *all* individuals. The goal is to determine classification boundaries among the *various* subjects. In ***authentication***, the population size is *unknown* and the training phase *only* gathers data from the user of interest. The goal is to determine the best authentication boundary for a *single* subject.

No study in the SOTA has tested its methods with both applications: they only focus on one, usually on identification, which is simpler than authentication. Our study analyzes both. Independently of the target application, achieving high biometric accuracy with PPG signals requires attaining a delicate balance between two competing goals:

- ***Challenge 1***: *reduce intra-cluster variance*. We need cardiac cycles that are as homogeneous as possible for the **same individual**, in order to obtain stable features.
- ***Challenge 2***: *increase inter-cluster distance*. We need cardiac cycles that are as different as possible **among individuals**, to define clear identification boundaries.

Figure 1 shows the PPG signals of two users collected in a controlled manner. These are the types of signals gathered in several SOTA studies [28, 36]. Under these favorable circumstances, it is simpler to tackle the above challenges and to differentiate the individuals.



(a) Controlled environment (b) Uncontrolled environment

Fig. 2. Cardiac periods collected for the same subject.

Morphologies. We use the term *morphology* to refer to the shape of a cardiac cycle, and *stable morphology* to refer to cardiac cycles that have (i) the same numbers of peaks and valleys, and (ii) a small signal variance. For example, subjects A and B in Figure 1 have stable morphologies with two peaks and three valleys. In uncontrolled environments, gathering distinct and stable morphologies for each user becomes significantly more complicated.

Metrics. There is no common metric in the SOTA to measure accuracy. Some studies use the EER [42], while others use F1-score [24] or BAC [28]. All these metrics are derived from true/false positive/negative results. Our evaluation can be presented with any of these metrics. We decided to use BAC because our datasets are unbalanced. BAC is the average of the true positive rate (TPR, sensitivity) and the true negative rate (TNR, specificity):

$$\text{BAC} = \frac{\text{TPR} + \text{TNR}}{2}, \tag{1}$$

where

$$\text{TPR} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}, \tag{2}$$

and

$$\text{TNR} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}. \tag{3}$$

3.3 The Detrimental Effect of Irregular Cycles

Multiple PPG studies report a high identification accuracy, ranging from 85% to almost 100%, depending on various evaluation parameters and scenarios [5, 15, 24, 28, 36, 40, 42]. Most of those studies, however, follow a *well-controlled* data-gathering process, which results in limited distortions across cardiac periods, and thus, a good performance. The *controlled* process is reflected in two factors: (1) the conditions under which the dataset is gathered, and (2) the diversity of individuals in the dataset. Controlled datasets typically focus on healthy individuals with a narrow age range, between 20 and 40, as discussed in Section 7.1. Furthermore, for each individual in the dataset, the measurements seem to be taken without considering the small but normal finger movements that affect the pressure between the fingertip and the sensor. This type of control group, consisting of young and healthy people without considering finger movements, leads to more stable signals.

In contrast to the *controlled* process, an uncontrolled process covers a more realistic scenario: a wider age range, including different health conditions, and considering minor (unconscious) finger and hand movements. Figure 2 depicts PPG signals for a single individual collected in controlled and uncontrolled environments. The small variance observed in Figure 2(a) is similar to the ones

Table 2. Performance of SOTA

	Controlled	Uncontrolled
Signal Variance	1.83	2.96
BAC for identification [24]	91%	72%
BAC for authentication [28]	93%	69%

observed in Figure 1 in [36] and Figure 3 in [28].[3] While it is not unreasonable to assume that PPG signals are collected in controlled environments, such assumptions constrain the ubiquitous applicability of PPG-based biometrics.

Differences in signal regularity can have a major impact on the performance of SOTA methods. Table 2 shows a preliminary evaluation with four subjects, for whom we collected PPG signals in controlled and uncontrolled environments. The exact description of the SOTA methods used as baselines for identification [24] and authentication [28], and the means used to calculate the signal variance, are explained in Section 7. For now, *the important takeaway is that when the SOTA is tested with controlled data, the performance is high (above 90%), as reported in the original studies; but when tested with highly variable signals, the accuracy drops significantly (around 70%).*

4 MORPHOLOGY STABILIZATION

A major shortcoming of the SOTA is to use the *same spectrum* to filter the PPG signals of *all subjects*. In this section, we propose a novel adaptive filtering method. Figure 3 depicts a macro view of our approach and its relation with the SOTA. First, we describe the methods we borrow from the SOTA (Section 4.1), and then, we describe their limitation and present our contributions (Section 4.2).

4.1 SOTA Methods: Basic Filtering and Derivatives

Figure 3(a) depicts an ideal PPG signal. The biometric signature of an individual is captured by four fiducial points: diastolic (highest valley), systolic (lowest peak), dicrotic notch (which forms a small peak in the middle of the period), and second wave. Figure 3(b) shows a raw PPG signal $s(t)$, which has two undesirable properties. First, a significant amount of noise distorts the location and intensity of the fiducial points, and in some cases, the noise level is high enough to erase the second wave and dicrotic notch completely, affecting the system's accuracy severely. Second, even in the ideal case, when all fiducial points are present, the signal's morphology is too simple and generic. Given that features are obtained based on the relative duration, heights, and slopes between fiducial points, the limited number of fiducial points limits the number of features. To overcome these effects, the SOTA proposes a basic filtering step and the use of the second derivative of the PPG signal.

Filtering. To mitigate the noise in PPG signals, the SOTA has identified the spectrum over which cardiac information is contained. For biometric purposes, the lowest meaningful frequency of a PPG signal is the heart rate. Considering that athletes can have heart rates as low as 0.5 Hz [45], the lower cut-off frequency f_l is usually set to that value. Regarding the upper cut-off frequency f_h, according to [10], sampling frequencies above 25 Hz do not provide any extra information, hence, f_h can be set to 12.5 Hz (due to the Nyquist–Shannon sampling theorem). Some studies use other filtering bands [24, 28], but the overall filtering process is similar. Figure 3(d) shows a PPG signal $f(t)$ after being filtered with a second-order Butterworth bandpass filter with a bandwidth of 0.5–12.5 Hz [6].

[3]These studies do not post their PPG data. To infer the variance of their signals, we have to rely on their figures.

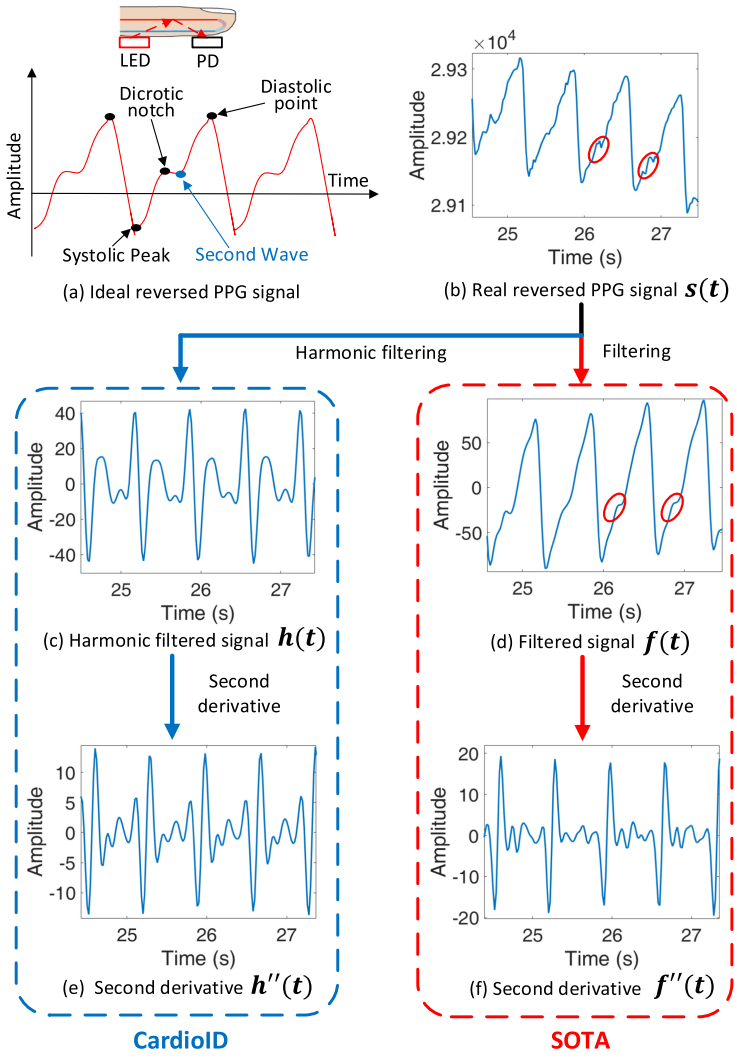
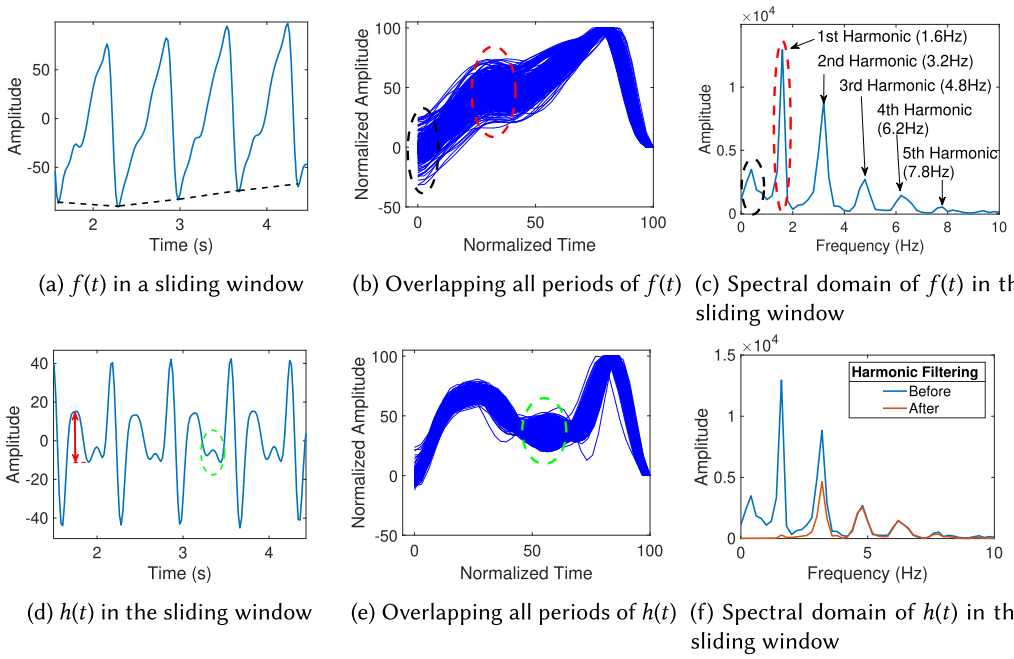


Fig. 3. Morphology stabilization.

Derivatives. Filtering alleviates noise, but it also eliminates valuable information. For instance, the raw PPG signal $s(t)$ in Figure 3(b) contains faint but detectable second waves (red circles). After filtering, however, those fiducial points no longer exist (corresponding red circles in Figure 3(d)). To overcome this issue, researchers obtain features not only from $f(t)$ but also from its second derivative $f''(t)$ [24]. Figure 3(f) depicts the second derivative of the filtered cardiac signal, which exhibits more fiducial points than $f(t)$.

4.2 Contribution: Harmonic Filtering

We also use the filtering and derivative stages, but we do not utilize the same parameters for all users. We propose a harmonic filtering phase that adapts its parameters to every individual. This process allows us to obtain more stable morphologies for every user (Challenge 1) and distinct fiducial points among users (Challenge 2). Our harmonic filtering depends on the subject's heart



(a) $f(t)$ in a sliding window (b) Overlapping all periods of $f(t)$ (c) Spectral domain of $f(t)$ in the sliding window

(d) $h(t)$ in the sliding window (e) Overlapping all periods of $h(t)$ (f) Spectral domain of $h(t)$ in the sliding window

Fig. 4. Frequency analysis to determine the lower cut-off frequency f_l.

rate, which can change over time, thus, we track the heart rate using a 5-second sliding window with 1-second steps.

4.2.1 Determining the Lower Cut-off Frequency f_l. The SOTA usually uses a lower cut-off frequency that is too low, which increases signal variance and makes it hard to identify the most vulnerable fiducial points (second wave and dicrotic notch). Figure 4(a) shows the filtered signal $f(t)$ using SOTA methods and Figure 4(b) shows the overlapping cardiac cycles using the endpoint of periods as an alignment anchor. We can observe a large variance in the starting points (black ellipsoid in Figure 4(b)) and significant instability in the dicrotic notch (red ellipsoid in Figure 4(b)).

Thus, *the fundamental question is how high should f_l be?* To obtain this optimal value, we analyze $f(t)$ in the spectral domain in Figure 4(c). Our analysis leads to two important insights. First, the wide variance occurs because an $f_l = 0.5$ Hz does not filter important dynamics such as heart rate variability, the effect of respiration (slow changing frequency component), and subtle unconscious pressure changes on the fingertip, which are common phenomena in uncontrolled scenarios. Those dynamics generate a fluctuating envelope in the time domain (black dashed line in Figure 4(a)), which causes the height differences between the starting and end points in periods. Considering that the endpoints are the alignment anchors, those height differences among periods will lead to a significant variance in the starting points. Second, an $f_l = 0.5$ Hz obscures the dicrotic notch. The energy of PPG signals is concentrated around the harmonics of the heartbeat, in particular the first harmonic (red ellipsoid in Figure 4(c)). The SOTA does not filter the first harmonic because it uses the heart rate period as a feature, which is good, but the spectral energy of the heart rate overwhelms the second wave and dicrotic notch, which are the most vulnerable fiducial points.

Our analysis indicates that to lessen the dampening effects of the heart rate period, we need to filter out the first harmonic. We noticed, however, that for some subjects the second harmonic is as high (and as dampening) as the first harmonic and should be attenuated too. Therefore, denoting

the frequency of the first harmonic as f_{1h}, we set $f_l = 2f_{1h}$. Note that with our approach we do *not* lose the heart rate data because it is contained in the other harmonics.

4.2.2 Determining the Upper Cut-off Frequency f_h. High-frequency noise modifies the location of fiducial points, which in turn, affects the stability of features and the overall performance of the system. Depending on the individual, a $f_h = 12.5$ Hz may be too high. For example, in Figure 4(c) the spectral energy is almost negligible beyond 10 Hz. Considering this situation, how low should f_h be?

As stated earlier, it is central to preserve the most vulnerable fiducial points on PPG signals (second wave and dicrotic notch). We use Figure 5, which zooms into those two vulnerable points, to illustrate the derivation of f_h. Denoting t_1 as the duration between the second wave and the dicrotic notch, the sine wave in the FFT containing the spectral energy of these points has a period of $2t_1$, which means that f_h must be higher than $1/2t_1$, else those two fiducial points would be filtered out. Now, denoting t_p as the period of a cardiac cycle, we observed empirically that $2t_1 > t_p/5$, and consequently, in the frequency domain $1/2t_1 < 5/t_p$. Finally, considering that $5/t_p$ represents the fifth harmonic of the heart rate, we set $f_h = 5.5f_{1h}$ to preserve all fiducial points while removing high-frequency noise. The negligible frequency components beyond the fifth harmonic in Figure 4(c) prove the correctness of our analysis.

4.2.3 Adaptive Filtering. The frequency response of our filter is solely based on the value of the first harmonic, $2f_{1h}$ to $5.5f_{1h}$, which is simple to obtain from the signals. More importantly, our approach is based on the subject's heartbeat instead of fixed parameters, allowing us to perform accurate adaptive filtering per subject. Figure 4(d)–(f) shows the signals filtered with our method, their overlapping cycles, and spectral domains. We can observe that, compared to the filtered signal $f(t)$ in the SOTA, $h(t)$ has three advantages: (i) the signal variance is much lower throughout the entire cycle, Figure 4(e); (ii) the difference between the second wave and dicrotic notch is accentuated significantly, red arrow in Figure 4(d); and (iii) our method exposes another fiducial point, green ellipsoid in Figure 4(d), which we can exploit to obtain more features as explained next.

To verify our choice of the frequency band, we designed two additional frequency bands for comparison: a wider band $[1.5f_{1h}, 6f_{1h}]$ and a narrower band $[2.5f_{1h}, 5f_{1h}]$. In this comparison, we feed the public dataset [38] (consisting of 35 subjects introduced in Section 7.1) to our whole authentication pipeline to check the performances of those harmonic filters (feature extraction in Section 5.3, authentication in Section 6.2). In the end, the average BAC results for the wider one, the narrower one, and ours are 92.03%, 91.89%, and 93.71%, respectively. Our selected band demonstrated a superiority of at least 1.5% compared to the other two. Therefore, our frequency band choice $[2f_{1h}, 5.5f_{1h}]$ is optimal.

4.2.4 Derivatives. As described earlier, the SOTA uses derivatives to accentuate the presence of fiducial points. We borrow that idea to obtain the second derivative of our harmonic signal $h(t)$. Figure 6 plots overlapping cycles for $f''(t)$ and $h''(t)$ for two sample subjects with uncontrolled data. Our second derivative $h''(t)$ has two important advantages compared to the SOTA's $f''(t)$. First, even though $f''(t)$ is more stable than $f(t)$ because the derivative removes offsets, $h''(t)$ is still less variable because it inherits the stability of $h(t)$. The variance of $f''(t)$ for subjects A and B are 2.8 and 3.0, respectively, while for $h''(t)$ are 2.2 and 2.8. This lower variability helps to tackle *Challenge 1*. Second, thanks to the tailored cut-off frequencies of our adaptive filter, $h''(t)$ can exploit the specificity of $h(t)$ to obtain more distinctive morphologies for different users, tackling *Challenge 2*. Compared to $f''(t)$, the fiducial points of $h''(t)$ are more distinctive and conspicuous across the *entire* time domain. Furthermore, subject A (blue) in Figure 6(b) shows that the second

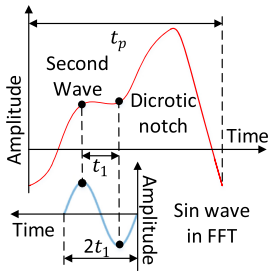
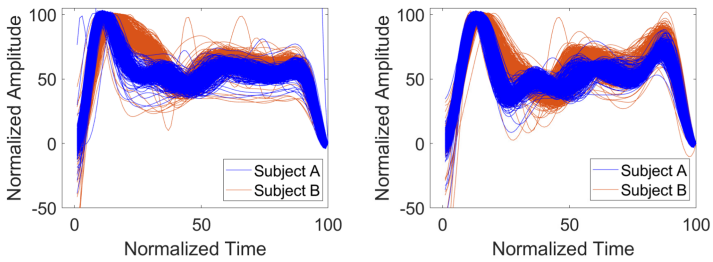


Fig. 5. Frequency analysis to determine the upper cut-off frequency f_h.

Fig. 5. Frequency analysis to determine the upper cut-off frequency f_h.



(a) SOTA: $f''(t)$ (b) CardioID: $h''(t)$

Fig. 6. Overlapping periods with uncontrolled data.

derivative disentangles the "knot" caused by the new fiducial point captured by the green ellipsoid in Figure 4(e).

Summary. Overall, our approach also follows the two basic steps of the SOTA, filtering and second derivatives, but using a novel filtering method leads to a more stable morphology for each user (Challenge 1) and more distinctive morphologies for different users (Challenge 2). The only input parameter required by our filter is the first harmonic (heart rate period), which can be easily obtained from any PPG signal. The SOTA obtains its features from $f(t)$ and $f''(t)$, and we obtain them from $h(t)$ and $h''(t)$. An exact description of the selected features is presented next.

5 MORPHOLOGY CLASSIFICATION

Existing studies share a common underlying assumption: all cardiac signals have a single dominant morphology. That, however, is not necessarily the case. We show that a single user can have multiple *valid* morphologies. Without this insight, a system would need to either discard periods that do not conform to a pre-defined morphology (introducing latency), or consider all periods with different morphologies, but at the risk of obtaining widely different features for the same user (reducing accuracy).

In this section, we first describe the segmentation method to obtain cardiac periods, then we show that cardiac periods can have multiple morphologies, and finally, we describe the features used in those various morphologies.

5.1 Signal Segmentation

Several methods can be used to segment periodic signals. Many approaches use amplitude-based thresholds to detect periodic peaks or valleys [14]; however, given the strong distortions present in our signals, we decided to use a spectral approach [22]. Considering that our harmonic filter $h(t)$ relies on the first harmonic f_{1h}, we design a segmentation method that also relies on f_{1h}. We use a filter with bandwidth $[0.5, 1.5] * f_{1h}$ to isolate the heartbeat period. A sample harmonic signal $h(t)$ and the corresponding signal used for segmentation $f_{1h}(t)$ are shown in Figure 7. In spite of the distortions, our approach can accurately map the valleys from $f_{1h}(t)$ to $h(t)$ and perform segmentation.

Our harmonic signal $h(t)$ can cope with movements and changes in finger pressure, but sometimes the movement of the finger is so strong that a period becomes invalid. Our segmentation method has the ability to discard those events. For example, for the signal in Figure 7, the first

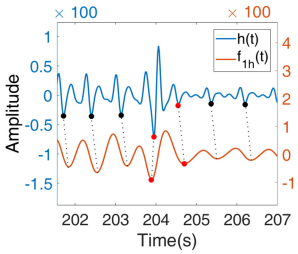
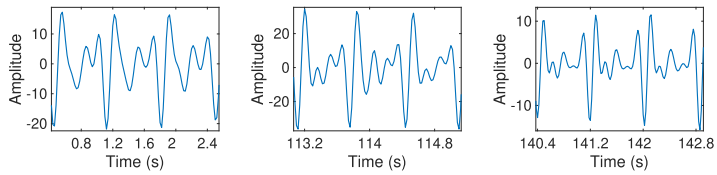
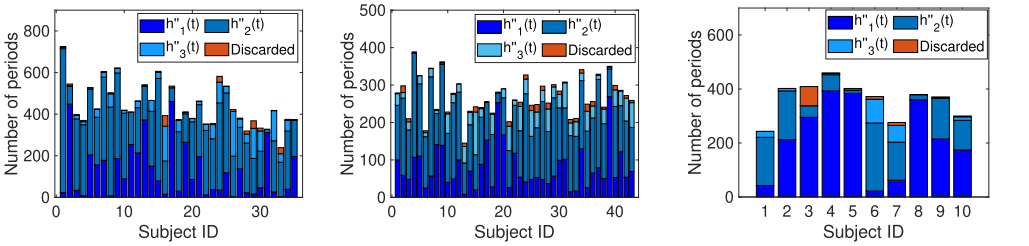


Fig. 7. Segmentation method.



(a) $h_1''(t)$: 3 peaks, 4 valleys (b) $h_2''(t)$: 4 peaks, 5 valleys (c) $h_3''(t)$: 5 peaks, 6 valleys

Fig. 8. Dominant morphologies for $h''(t)$.



(a) Public (pulse oximeter) (b) Private (camera) (c) Public (pulse oximeter, MIMIC-III)

Fig. 9. Frequency of occurrence of morphologies in three datasets. The datasets use pulse oximeters and cameras.

three and the last three periods (black dots) are valid, even if the finger pressure is different, but the middle period (red dots) captures a drastic finger movement that should be invalid. We introduce two criteria to verify a period. First, on the x-axis, the interval between two adjacent valleys on $h(t)$ must be similar to the period corresponding to the heartbeat. Second, on the y-axis, the values of the valleys at the start and end of a period should be similar. For the signal in Figure 7, the period with the red dots is discarded because it violates the second criterion. The segments for $h''(t)$ are obtained following a similar mapping approach.

5.2 Multiple Morphologies

Currently, all studies using fiducial points assume a single macro morphology for all subjects. That is a valid approach in controlled scenarios, but in uncontrolled scenarios, various factors can cause the appearance of multiple morphologies: unintended fingertip pressure [9], significant differences in the cardiac profiles of subjects, and so on. When we perform the second derivative of our harmonic signal $h(t)$, we observe multiple morphologies. Figure 8 depicts the three most dominant macro morphologies observed in $h''(t)$: $h_1''(t)$, $h_2''(t)$, $h_3''(t)$.

Our evaluation considers three datasets, and those dominant morphologies account for (i) 98.4% of the periods measured in a *public dataset* with 35 subjects [38], 15,301 out of 15,557 periods; (ii) 97.5% of the periods measured in a *private dataset* with 43 subjects that we collected from volunteers, 11,328 out of 11,617 periods; and (iii) 97.2% of the periods measured from 10 subjects in ICU (intensive care unit) from the public *MIMIC-III dataset [20]*, 3,509 out of 3,612 periods. Figure 9 shows the presence of the three macro morphologies in those datasets. There are other macro morphologies in $h''(t)$, but we do not consider them because they are rarely present ("Discarded" label in the figures).

If we would choose only one morphology as the template for all subjects, as conventional methods do, the system could face two major problems. *First*, it may take a long time to identify a subject

Table 3. Features Extracted from $h(t)$ and $h''(t)$

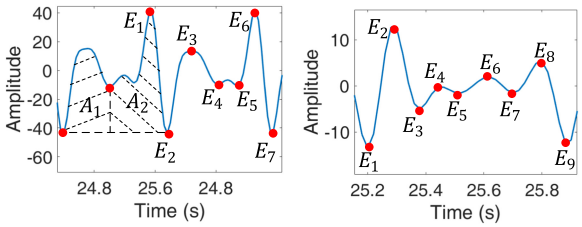
	Feature	Description
$h(t)$	Period duration	$E_6(t) - E_1(t)$
	Area ratio	A_1/A_2
	Duration (D_i)	$E_{i+1}(t) - E_i(t), i = 2, 3, 5, 6$
	Height (H_i)	$\|E_{i+1}(a) - E_i(a)\|, i = 2, 3, 5, 6$
	Slope	$(-1)^{i+1} * H_i/D_i$
$h''(t)$	Duration (D_i'')	$E_{i+1}''(t) - E_i''(t), 1 \leq i \leq 8$
	Height (H_i'')	$\|E_{i+1}''(a) - E_i''(a)\|, 1 \leq i \leq 8$
	Slope	$(-1)^{i+1} * H_i''/D_i''$

because the system will need to wait for the right morphology to arrive. For example, for the public dataset, our system can obtain 1.229 (15,301/12,444) periods/s, compared to much lower speeds if would only use $h_1''(t)$ (0.3584 periods/s), $h_2''(t)$ (0.7427 periods/s), or $h_3''(t)$ (0.1283 periods/s). *Second*, and perhaps more critical, the right morphology may never arrive for some of the subjects or it may be so rare that there would be insufficient samples to train the system properly. In practice, such a limitation would render an identification system futile because the basic premise is that it should be able to identify *all* members in a target group. In uncontrolled scenarios, no morphology is dominant. Even $h_2''(t)$, which is the most common, may be rarely active in some subjects, such as user 31 in the public dataset, and subjects 5 and 8 in the MIMIC-III dataset. Hence, the key advantages of considering multiple morphologies are decreasing latency and eliminating the risk of excluding some types of subjects.

5.3 Feature Extraction

We extract features from $h(t)$ and $h''(t)$. Following the approach employed in previous studies conducted in this field [24, 27, 28], our feature extraction process predominantly relies on capturing the geometric relationships among fiducial points. Compared with spectral features which aggregate multiple periods as an authentication unit, our fiducial features support authentication on each period. This distinction contributes to a reduction in authentication latency, facilitating the real-time operation of our system. Figure 10 and Table 3 provide a pictorial representation and the notation for all the features. In our notation, $E_i(t)$ and $E_i(a)$ denote the time and amplitude of fiducial point i. For $h(t)$, shown in Figure 10(a), we collect three types of features: (1) the duration of a period, (2) the ratio of the areas inside a period, and (3) the differences in duration, height, and slope between consecutive fiducial points in one period. The total number of features for $h(t)$ is 14. For $h''(t)$, we only consider the third type of feature (differences between contiguous fiducial points). Figure 10(b) shows the features for $h_2''(t)$, and the same principle is applied to extract the features from $h_1''(t)$ and $h_3''(t)$. In the end, the number of features for $h_1''(t)$, $h_2''(t)$, and $h_3''(t)$, are 18, 24, and 30, since they have 3, 4, and 5 peaks, respectively (Figure 8). Features based on duration and height are susceptible to heartbeat variance. In [19, 28], the authors state that normalizing the features makes them immune to heart rate changes. Therefore, we also normalized the duration and height of $h(t)$ and $h''(t)$.[4]

[4]For $h(t)$ we also observed two types of morphologies: one where E_4 and E_5 are present and the other where those two points merge into a single valley. However, contrary to the multi-morphology approach used for $h''(t)$ in Section 5.2, we decide to use a single morphology for $h(t)$ because the features obtained from the fiducial points E_4 and E_5 did not have any impact on the system's accuracy. The information from those two points gets disentangled and captured in one of the three morphologies present in $h''(t)$. Due to this reason, we do not evaluate $i = 4$ for $h(t)$ in Table 3.



(a) Fiducial points on $h(t)$ (b) Fiducial points on $h_2''(t)$

Fig. 10. Fiducial points used to extract features.

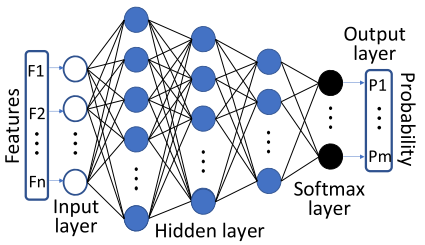


Fig. 11. The structure of our NN.

6 IDENTIFICATION AND AUTHENTICATION

As stated earlier, SOTA studies only evaluate one type of application: identification or authentication (mainly identification). We consider both. Our system relies on the same set of features for both cases. Upon receiving a raw PPG period, we first obtain $h(t)$ and $h_i''(t)$, and derive their features. The combined features, $h(t)$+$h_i''(t)$, are given as inputs to two different processing branches depending on the type of application. Considering that performing identification is simpler, we first present that system, and later we focus on authentication.

6.1 Identification

Identification requires gathering training data from *all subjects*, and during the testing phase the aim is to match an incoming cardiac sample to the right subject. As with many other classification problems, PPG identification requires two main components: *dimensionality reduction*, to identify the most informative features; and *decision boundaries*, to perform accurate classification.

The SOTA utilizes two supervised learning methods, linear and non-linear, but does not provide insights about which one is better and why. In our evaluation, we consider both methods. The most representative linear method is **linear discriminant analysis (LDA)** [46], which simultaneously reduces dimensionality and draws decision boundaries. The most representative non-linear methods are based on **neural networks (NNs)** [42]. As it is customary with NN [18], we first use an autoencoder for dimensionality reduction, blue layers in Figure 11, and then, we add a softmax layer to perform classification (decision boundaries), black layer.

Considering that we use three morphologies, we need three LDA and NN pipelines running in parallel for each morphology (each pipeline receives the corresponding set of features presented in Table 3). Since LDA is an analytical solution, the LDA module is the same for all three pipelines (but with different training data). In contrast to LDA, due to the influence of the network structure and parameter values, we tailor three different NN modules for each morphology. The hidden (blue) layers for morphologies one, two, and three are 128-64-32, 170-85-42, and 128-64-32, respectively. The activation functions are sigmoids to guarantee the non-linearity of the system, and parameters such as L2 and sparsity regularization are tuned for each morphology.

6.2 Authentication

Contrary to identification, in authentication systems, the training set only consists of samples from the *legitimate subject*, while its testing set can include samples from *any subject*. Authentication also requires dimensionality reduction and boundaries, but given that we lack information about other users, drawing an optimal boundary for that single legitimate user becomes more complex. Next, we first explain the methods used in the SOTA for dimensionality reduction, and then, some techniques to improve the definition of boundaries.

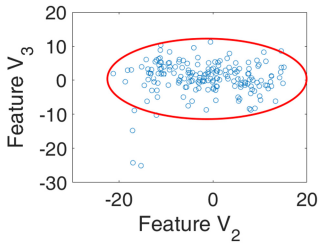


Fig. 12. Mahalanobis.

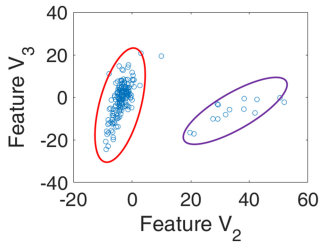


Fig. 13. Multi-cluster.

Dimensionality reduction can be performed with linear and non-linear methods. There are two mainstream linear techniques: **principal component analysis** (**PCA**) [32] and **non-negative matrix factorization** (**NMF**) [26]. NMF requires non-negative features, but the slopes in our feature set can be negative. Hence, similar to prior studies [28], we adopt PCA. Even though there are several non-linear dimensionality reduction techniques—such as Isomap [41], **local linear embedding** (**LLE**) [34], **t-distributed stochastic neighbor embedding** (**t-SNE**) [29], and autoencoder [18]—we did not find SOTA studies using them for PPG authentication. Isomap, LLE, and t-SNE share a common disadvantage for PPG authentication: they must perform an entire recalculation every time a new test point is added. Autoencoders, on the other hand, do not have that shortcoming. We performed a preliminary evaluation of authentication with autoencoders but the performance was not good. We hypothesize that it is due to the limited data, autoencoders are usually trained with at least thousands of training points.[5] PPG-based systems are trained with a few minutes of cardiac data in one subject, which maps to a few hundred cardiac periods. In identification, NN methods can be trained with several thousand samples coming from *all users*, but in authentication, we only have a few hundred samples coming from *the legitimate user*. Due to this finding, in our evaluation section, we only consider PCA for authentication.

Mahalanobis distance. After dimensionality reduction, the most significant features of a subject usually form a cluster similar to the one shown in Figure 12. When a new test sample arrives, the system calculates the average distance of this new point to the cluster. If the distance is below a threshold, the user is deemed legitimate. Many studies use Euclidean distances to measure proximity [28]. But Euclidean distances are fundamentally ill-equipped to deal with feature spaces that have widely different variances. For example, in Figure 12, using Euclidean distances, with any threshold, leads to a boundary that has the shape of a circle. The circle will be either too long for feature v_3, causing numerous false positives; or too short for v_2, causing significant false negatives. Therefore, we adopt the Mahalanobis distance [31], which considers the standard deviations in each dimension and can be used to define tight boundaries such as the red ellipsoid shown in Figure 12.

Multi-cluster approach. Current PPG authentication systems assume that the features of a user converge to a *single* cluster [28, 46]. However, with uncontrolled data, we observed that a single subject can form two or more clusters for a single morphology, as depicted in Figure 13. We need an authentication system that can identify multiple clusters and then use the Mahalanobis distance to set an appropriate threshold for each cluster.[6]

[5] At a popular Quora forum discussing "How large should be the dataset for training a Deep autoencoder?" Yoshua Bengio states the need for having large amounts of training data [3].

[6] To calculate the Mahalanobis distance, the number of samples must be greater than the number of features (dimensions). If a cluster has few samples, like the purple one in Figure 13, we use spline interpolation to add the extra necessary points.

Table 4. Details of the Three Datasets Used in Our Evaluation

	# subjects	# female	Age (Mean, Variance, Range)	RD (mins)	CTE
Public (pulse oximeter)	35	12	28.4, 14.04, 10–75 Y/O	5–6	4.29
Private (camera)	43	16	36.7, 14.93, 12–79 Y/O	4	5.97
MIMIC-III	10	8	Encrypted	5	4.72

For our purposes, the clustering method should meet three requirements: (i) be resilient to the presence of outliers, (ii) able to detect clusters with arbitrary shapes, and (iii) fast. Hierarchical clustering methods, such as BIRCH [33], and centroid-based methods like K-means [30] are vulnerable to outliers and cannot detect arbitrary shapes. Most grid-based clustering methods, like CLIQUE [1], and density-based methods, like OPTICS [2] and DBSCAN [13] do not have shortcomings (i) and (ii), but they need a relatively long computation time. Due to the above reasons, we decided to use WaveCluster [37], which exploits the multi-resolution property of wavelet transforms. WaveCluster can identify arbitrary shape clusters at different degrees of accuracy, it is insensitive to outliers and has a low time complexity $O(n)$.

7 PERFORMANCE EVALUATION

In this section, we describe the datasets we use, the studies taken from the SOTA as baselines for comparison, and the results for the evaluation of identification and authentication.

7.1 Datasets

We use three datasets to evaluate the performance of CardioID. The first dataset uses a *pulse oximeter* and is public [38]. All subjects are sitting during the signal collection. The second dataset is collected by us from volunteers who are also sitting.[7] We use the *camera and flashlight* of an iPhone-7 to record 60-FPS videos of the volunteers' fingertips. In each frame, we focus only on the red channel of the pixels covered by the fingertip. The method to select the covered pixels is the same as in [28]: $I_{\text{red}} > 80\% \times (I_{\text{red}} + I_{\text{blue}} + I_{\text{green}})$. We average the red-channel intensities among the selected pixels to represent one data point of a PPG signal. To maximize the peak-to-peak amplitude of cardiac periods, we carefully set the three parameters affecting the camera's exposure: the aperture and ISO are set to the lowest values, −2 and 20, respectively, and the shutter speed to 200. Other parameters like white balance, focus, and zoom are set to auto. The third dataset is obtained from patients in the MIMIC-III waveform database [20]. All subjects from this dataset were ICU patients (intense care unit) at a Medical Center in Boston, USA. Their PPG signals are gathered with a pulse oximeter in multiple sessions. We select 10 patients from that dataset and their information is shown in Table 5.

Significance of datasets. The parameters of the three datasets—the number of subjects, gender, age distribution (average, variance, and range), the **recording duration** (**RD**), and variability (**cross-track error** (**CTE**))—are given in Table 4. There are three important points to highlight about the selection of our datasets.

First, *no SOTA study has analyzed the performance of their methods using both types of sensors, pulse oximeter, and camera.* In general, a pulse oximeter is more precise than a camera because its infrared spectrum can enhance the signal quality, and its finger clip can reduce noisy motion artifacts [14]. This is one reason why the CTE (variability) in Table 4 is higher for the camera dataset.

Second, *our datasets consider a more diverse group of people.* Even for *healthy* people, which is the main focus of the SOTA, the morphology of cardiac signals can vary significantly based on the

[7]This dataset was gathered under the approval of the Ethics Committee (HERC) of our university.

Table 5. Information of the Subjects in the MIMIC-III Dataset

Subject ID	Gender	Ethnicity	Training data collection time	Testing data collection time	Interval (hours)	Disease
10045	F	W	23:32	9:31	10	Fever
10083	F	W	12:28	23:28	11	Hypetension
10124	F	Jewish	9:34	18:05	9	Congestive heart failure
41976	M	Hispan	23:18	20:50	21	Pneumonia
42033	F	W	17:29	13:38	20	Shortness of breath
42199	F	W	12:41	20:09	8	Chest pain
42302	F	W	12:42	19:32	7	Asthma COPD Flare
42367	F	W	11:59	19:59	8	Seizure; status epilepticus
43798	M	W	18:27	5:34	11	Esophageal CA/SDA
43827	F	Unknow	13:29	20:47	7	MI congestive heart failure

age group and skin tone. This is another reason why the CTE in Table 4 is higher for the camera dataset. The studies we use as baselines, [28] and [24], focus on a narrow age segment of the adult population, 22–33 and 18–46, respectively. The age ranges of the first two datasets we use are 10–75 (public) and 12–79 (private), including children, teenagers, adults, and elders. Moreover, we also consider a bigger population: 70% more for the camera sensor (43 people vs. 25, [28]) and 17% more for the pulse oximeter (35 vs. 30, [24]). In terms of skin tones, the public dataset includes only medium-toned skin, and our private dataset includes 51% medium, 33% light, and 16% dark skin.

Third, *no study considers unhealthy individuals or long periods of time between the training and testing periods*. To identify the limits of PPG-based identification and authentication, we consider a dataset (MIMIC-III) with people having various health-related problems from asthma and seizures to heart failures and hypertension (Table 5). For these subjects, we select signals for the training and testing phases that are **at least 7 hours apart**. Since these individuals are ICU patients, they are subject to the effects of medicine, which can change their physiological and psychological conditions even within a session.

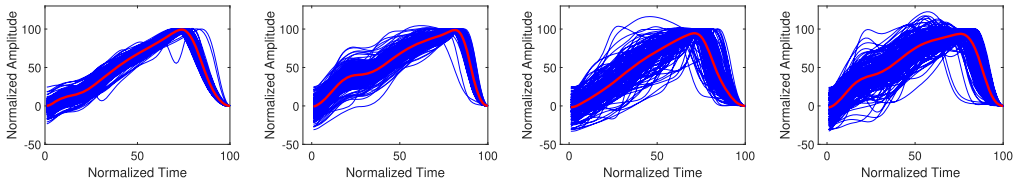
Overall, to the best of our knowledge, these datasets consider—by a wide margin—the most demanding conditions in the SOTA.

Signal variance analysis. Gathering data from different sensors, while considering motion artifacts and a broader range of people, provides us with more realistic (less controlled) PPG signals. To quantify the variance of these signals, we first obtain the average signal for a user, red signals in Figure 14; and then, we calculate the CTE[8] from every (blue) PPG signal to its average. Denoting e_i as the CTE for signal i, the signal variance for a subject is the mean absolute error for all e_i's. The average signal variances for our datasets are 4.29, 5.97, and 4.72, as shown in Table 4. *In order to put these values in context, it is important to note that the variance found in SOTA plots is a bit lower than what is shown in Figure 14(a), i.e., less than 2.* The majority of users in the public dataset have a variance in the range $[2, 6]$; in the private dataset in $[4, 6]$, and in the MIMIC-III dataset in $[2, 6]$. Hence, our evaluation copes with a wide spectrum of signal variability.

7.2 Baselines Used for Comparison

We utilize two SOTA studies as baselines for comparison, one for identification [24] and the other for authentication [28]. The reasons for selecting those baselines are presented in Section 2. In this subsection, we quantify the improvement in the acquisition rate for CardioID and the difference in accuracy between our work and the SOTA baselines.

[8]The CTE is used in GPS systems to measure the difference between the given and followed paths. We tried different similarity metrics, including dynamic time warping, and we found that the CTE captures the similarity of PPG signals in a more precise manner.



(a) Variance 1.8 (b) Variance 3.9 (c) Variance 5.8 (d) Variance 6.6

Fig. 14. Different signal variances.

7.2.1 Quantifying Acquisition Rates. Every study in this research area, including ours, removes periods that do not conform to the required morphologies. The goal is to discard as few periods as possible while maintaining high accuracy. Denoting S as the cardiac periods from all users and S' as the periods used by a system (after discarding non-conforming morphologies), the acquisition rate is given by S'/S. With controlled data, the acquisition rate is high, $S' \approx S$; but with uncontrolled data, the rate can be very low, $S' \ll S$. As stated in Section 5, a low rate can increase the system's delay and in some cases exclude the participation of some users. Hence, before even assessing the accuracy of the system, we need to make sure that a method has the capability to recognize 100% of the users.

Considering that S is equal to 14,347, 10,728, and 3,612 periods for the public, private, and MIMIC-III datasets, respectively, we first need to find S' for the SOTA baselines. The morphology and features used by Kavsaouglu et al. are presented in Figures 7 and 8 in [24], and the corresponding information for CardioCam is provided in Figure 8 in [28].[9] We use that information to discard the morphologies that do not conform to their requirements because the right morphology is necessary to obtain their features. After discarding the non-conforming morphologies in the SOTA, we obtain the following acquisition rates: 74.6% (public dataset), 64.5% (private dataset), and 66.3% (MIMIC-III dataset) for [24]; and 59.2% (public), 32.8% (private), and 37.4% (MIMIC-III) for [28]; significantly lower than the 98.4% (private), 97.5% (public), and 97.1% (MIMIC-III) obtained for CardioID. *More importantly, in the camera dataset, there were three users that* ***did not have a single cardiac period*** *resembling the morphology required by [28], and thus, there is no possibility to authenticate them with that method.*

Moreover, the SOTA studies [24, 28] show a long acquisition delay. There are 12,444, 10,320, and 3,000 seconds of data for the public, private, and MIMIC-III, respectively. The acquisition speeds are 0.932 (public), 0.726 (private), and 0.798 (MIMIC-III) periods/second for [24]; and 0.740 (public), 0.369 (private) and 0.450 (MIMIC-III) periods/second for [28]; significantly lower than the 1.229 (private), 1.097 (public) and 1.169 (MIMIC-III) periods/second obtained for CardioID. The SOTA acquisition speeds are below 1 period/second. Some of the acquisition speeds are even lower than 0.5 period/second, which means that users would need to wait a few seconds to be identified.

7.2.2 CardioID Variants. To assess the impact of our contributions—*morphology stabilization*, *morphology classification*, and *the reduction of non-linear effects for authentication*—we perform an ablation study creating different variants for CardioID. For the SOTA approaches, we implemented them based on the morphologies and features provided in their respective studies.

For identification, we consider four variants.

— Variant I.1 (MS): we use **morphology stabilization** to obtain $h(t)$ and $h''(t)$ with their respective features, cf. Table 3. This variant only considers morphology-2 periods, $h''_2(t)$. The classification method is K-NN, the same as in [24].

[9]Upon close inspection, we notice that, in both SOTA studies, the second derivative of $f(t)$ is a signal similar, but not exactly the same, as morphology-2 in our case, $h''_2(t)$.

Table 6. Percentage of Detectable Subjects and Periods in the Public (Pulse-oximeter) Dataset

Public	Signal variance	2	4	6	All
[24]	Subject %	82.9	100	100	100
	Period %	21.6	54.2	67.3	74.6
[28]	Subject %	68.6	80	82.9	82.9
	Period %	19.5	45.9	55.2	59.2
MS	Subject %	57.1	97.1	97.1	97.1
	Period %	8.1	29.8	44.9	59.4
MC	Subject %	74.3	100	100	100
	Period %	14.8	50.6	75.1	98.39

— Variant I.2 (MC): we add **morphology classification** to the MS variant. Here, we include periods with morphologies $h_1''(t)$ and $h_3''(t)$ but we still use K-NN for the classification.
— Variants I.3 and I.4 (CardioID.LDA and CardioID.NN): we replace K-NN in the MC variant with **LDA** and **NN**, respectively. We consider these variants the final implementations of CardioID for identification.

For authentication, we also consider four variants.

— Variant A.1 (MS): similar to the MS variant used for identification, but instead of K-NN, it uses PCA and Euclidean distance to achieve authentication, as in [28].
— Variant A.2 (MC): it adds **morphology classification** to the MS variant.
— Variant A.3 (Mahal): it replaces the Euclidean distance with **Mahalanobis distance** in the MC variant.
— Variant A.4 (CardioID): it adds the **multi-cluster** approach to the Mahal variant.

7.2.3 Emulating a Wide Range of Signal Variances. Our aim is to evaluate the SOTA baselines and CardioID variants under a wide range of signal variances. Collecting that type of data would require asking users to steadily increase the level of finger movement and pressure from low to high. That would be a complex process, instead, we use our datasets to create (emulate) subsets with increasing levels of variance. To generate the emulated subsets, we perform the following process. For every user, we only include signals that lead to a variance less than t, where $t = 2, 4, 6$. If after this filtering process, a variant cannot collect 20 periods from a user, we leave the user out of the emulated set because we would not have sufficient training data for that user. A macro view of the emulated subsets is presented in Tables 6–8. For example, if we look at Table 6 for reference [24], we can see that if we set $t = 2$, (i) only 21.6% of the periods would have a variance less than 2 **and** satisfy the morphology requirement of [24]; and (ii) only 29 subjects, out of 35 (82.9%), have more than 20 periods satisfying the conditions in (i).

Tables 6 and 7 provide two important insights. First, for all variance levels, the MC variant has the best performance in terms of including more users and having the highest acquisition rate (because it accepts three different morphologies). The SOTA baselines and the MS variant have lower performance because they consider only one morphology. Second, when we consider all the data (last column in the tables), one of the baselines [28] cannot include all users in either dataset. This is an important point because it means that the (only) morphology allowed by [28] has requirements that are so stringent that some users may rarely (or never) show the required morphology. *In fact, with the camera dataset, three users did not have a single cardiac period satisfying the morphology requirements, and two of those users were above 50 years old (an age bracket that was not considered in the SOTA).* The broad type of users in our dataset enables us to expose this type of

Table 7. Percentage of Detectable Subjects and Periods in the Private (Camera) Dataset

Private	Signal variance	2	4	6	All
[24]	Subject %	37.2	90.7	100	100
	Period %	5.5	33.1	51.2	64.5
[28]	Subject %	25.6	41.9	58.1	69.8
	Period %	4.7	18	25.9	32.8
MS	Subject %	18.6	81.4	95.3	100
	Period %	3.4	22.3	37.5	55.5
MC	Subject %	30.2	93	100	100
	Period %	5.5	37.6	64.3	97.5

Table 8. Percentage of Detectable Subjects and Periods in the MIMIC-III Dataset

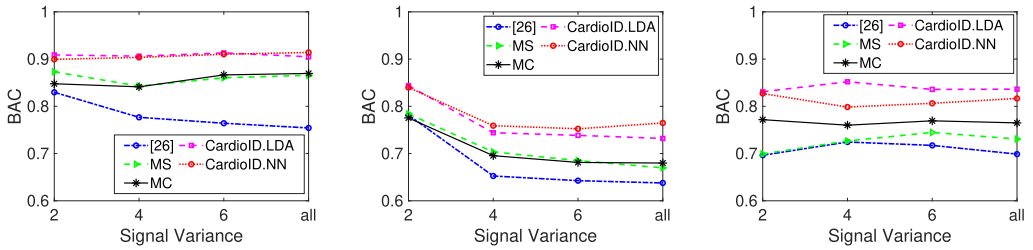
MIMIC-III	Signal variance	2	4	6	All
[24]	Subject %	70	90	90	90
	Period %	28.4	53.1	60.6	66.3
[28]	Subject %	40	80	80	80
	Period %	11	29.3	34.2	37.4
MS	Subject %	70	80	90	90
	Period %	8.6	20.7	26.3	31.8
MC	Subject %	80	100	100	100
	Period %	25.2	59.8	77.9	97.1

limitation. The variants LDA, NN, Mahal, and Cluster, have the same % of subjects and periods as MC because they are derived from that variant.

7.2.4 Providing the Right Context for Performance Evaluation. Until now, we have only evaluated the impact of CardioID on the acquisition rate. In the remainder of this section, we evaluate its accuracy, and it is important to consider the following points: (i) for each emulated subset, we use the first 80% of periods for training and the rest for testing; (ii) the variance observed in the SOTA signals is around 1.5, hence, the variance considered in our emulated subsets ($t = 2, 4, 6$) poses a greater challenge; (iii) for CardioID and the SOTA baselines, we use a *single period* to perform identification or authentication, using more periods would increase the performance but also latency; (iv) using a single cardiac period with *controlled signals*, the SOTA reports a BAC of 0.95 [28], which can be translated, in expectation, to the rightful user having a probability around 95% to get access to the system (sensitivity), and an attacker a probability of around 5% of being successful (specificity). Thus, our goal is to try to get as close as possible to a BAC of 0.95 with *uncontrolled signals*. In authentication and identification, *improvements in the order of 5%, or above, are considered significant.*

7.3 Identification

Public dataset. Figure 15(a) shows the results for the pulse oximeter dataset. The MS variant provides the most significant improvement, above 10% for most approaches. This occurs because our harmonic filter adapts to every user, enabling more distinct and stable features. The MC variant does not really improve the BAC, but—as described before—the fact that it accepts multiple morphologies improves inclusion (more subjects are accepted) and the acquisition rate (40% higher



(a) Public (pulse oximeter) (b) Private (Camera) (c) MIMIC-III (pulse oximeter)

Fig. 15. Identification performance.

than the MS variant and 20% higher than the SOTA baseline [24] for the entire dataset, cf. last column in Table 6). As for the *LDA* and *NN* variants, the performances are similar. The improvement of both solutions over the MC variant is around 5%. The final BAC of CardioID reaches 91.6%, close to the 95% target, while the SOTA drops to 76.1% [24], leading to a total improvement of around 15%.

Private dataset. Figure 15(b) shows the results for our camera dataset. We can see that, compared to the prior dataset, the overall performance is lower, but the same general trends appear, showcasing the ability of CardioID to improve the identification performance across different sensors and users. Considering the subsets with variances four and above, we obtain the same benefits as for the pulse oximeter: the MS variant provides about a 4% improvement, the MC variant offers a marginal improvement in terms of accuracy but significant improvements on subject inclusion and period acquisition rate (cf. Table 7), and the classification methods (LDA and NN) provide an improvement of around 7%.

It is important to note that even though CardioID performs significantly better than the SOTA with the camera dataset (12% better), it is still far from the 95% BAC target. In expectation, an 80% BAC is not reliable because it gives the right user an 80% chance to access the system and attackers a 20% chance. At the end of this section we discuss some ways to overcome this problem.

MIMIC-III dataset. Figure 15(c) shows the result for the patients in ICU. The cardiac signal is obtained with a pulse oximeter, but *the time difference between the training and testing for all subjects is at least 7 hours. In the prior two datasets, as well as for all the other SOTA studies, the training and testing sets are contiguous.* The overall performance of the MIMIC-III dataset is better than with the camera dataset. The difference between the SOTA baseline [24] and the MS variant is about 3%. The MC variant, whose multiple morphologies allow to nearly triple the number of cardiac periods and consider 10% more subjects (cf. Table 8), improves the performance by 3% with respect to the MS variant. Although the NN classification method is inferior to LDA, it still provides a 5% improvement. The LDA method, which is the one that performs best, can obtain a BAC of 83%, which is a 15% improvement compared with the SOTA [24].

Linear vs. non-linear methods. SOTA studies have been using linear [5, 36, 46] and non-linear [40, 42] methods to perform identification, but no study has compared both approaches or stated why one is preferable over the other. Our evaluation shows that both methods have a similar performance (Figure 15). We hypothesize that this occurs because our morphology stabilization and classification provide cardiac periods with stable and distinct features, and hence, the role of the classification method is less prominent. To highlight the importance of our morphology variants (MS and MC), we replace the K-NN classifier used in [24] with LDA in the camera dataset (while maintaining everything else the same). The results in Figure 16 show that LDA even degrades the performance a bit. Without the stable and distinct PPG signals provided by our morphology stabilization and classification, a machine-learning method cannot do much on its own to overcome the high variance present in uncontrolled scenarios.

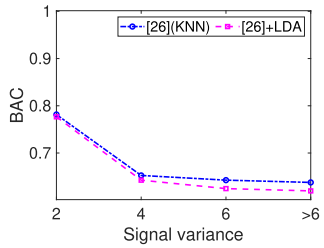


Fig. 16. LDA contribution.

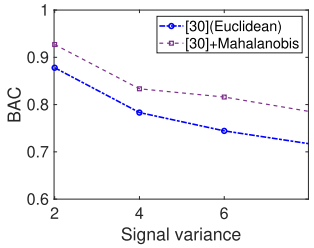
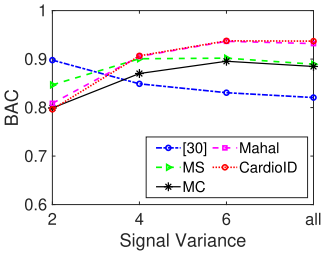
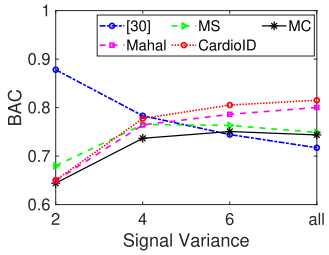


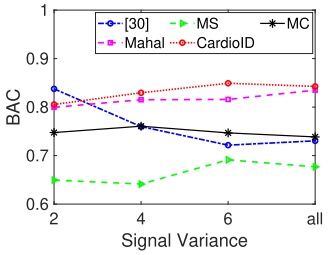
Fig. 17. Mahalanobis contribution.



(a) Public (pulse oximeter)



(b) Private (camera)



(c) MIMIC-III (pulse oximeter)

Fig. 18. Authentication performance.

7.4 Authentication

Public dataset. Before discussing the authentication results, we need to highlight a critical difference compared to identification: for authentication, the SOTA [28] **fails** to include all types of users. In Table 6, we can observe that both, the SOTA baseline used for identification [24] and the MC variant, consider all 35 subjects for $t = 4$ and above, and thus, the comparison is unbiased because the population size is the same. However, *the maximum number of subjects considered by [28] (baseline for authentication), is only 29 (82.9%).* This occurs because our evaluation requires at least 20 periods per subject, but the morphological requirements of [28] are too stringent, which does not allow getting enough periods for six subjects. Hence, for the results we present with authentication, CardioID faces a more challenging scenario than the SOTA because a bigger population increases the likelihood of errors.

With that clarification, we can now discuss the main insights for the private dataset (pulse oximeter) in Figure 18(a). First, at $t = 2$, all the approaches have a similar performance. This occurs due to the limited data. For $t = 2$, the emulated subset filters out most samples. Contrary to identification, where the system can exploit the samples from *all the other users*, in authentication, the system can only use the limited samples belonging to a *single* user. Thus, for $t = 2$, the performance of the system is largely determined by the small number of relatively well controlled signals, leaving little room for the methods to showcase their respective strengths. For $t = 4$ and above, the MS and MC variants play the same role as in identification: MS increases the performance, while MC increases the participation (number of subjects) and the acquisition rate (reduces delay). Overall, CardioID achieves a 93.7% BAC with 35 subjects, while [28] achieves 11% less BAC with only 29 subjects.

Private dataset. Figure 18(b) shows the results with the camera dataset. Due to the higher signal variability, the SOTA baseline filters even more periods than with the prior dataset: the percentage of subjects for the SOTA is less in Table 7 than in Table 6 (69.8% vs. 82.9%), and as stated before, three subjects did not have a single period satisfying the morphological requirements

in [28], which would make the SOTA system invalid for this target group. Even if we leave that critical point aside, CardioID still outperforms the SOTA by 10%, but it is still not able to reach the desired 95% BAC target. A counter-intuitive trend with the camera dataset is that CardioID's performance increases significantly with the signal variance. But this is not due to the increase in variance per se (which adds noise), but due to the increase in data (when more variability is allowed, the subset contains more cardiac periods and subjects).

There is one result, however, that we did not expect and it highlights a particular strength of the SOTA with the camera. For $t = 2$, the baseline [28] has a strong performance compared to all of our variants, and it is in accordance with what the authors report in the original article.[10] Initially, we thought that it was because, at $t = 2$, the SOTA considers only 11 out of 43 subjects (25.6% subject participation in Table 7), but our MS variant, which relies only on morphology-2, also considers the same amount of subjects and performs worse. The SOTA has strict morphological requirements that filter out many cardiac periods, but this conservative standard allows them to have more similarity among their periods when the data is controlled, which is particularly useful in authentication because the training phase utilizes a single user. The stability of our features relies solely on the harmonic filter. After that, the conditions to consider a morphology valid are rather permissive: simply counting the number of peaks and valleys present.

MIMIC-III dataset. Figure 18(c) shows the results for the ICU patients. Similar to the other datasets, the benchmark [28] cannot include all subjects (80%). The performance of the MS variant is noticeably below other methods, due to the limited amount of data (cf. Table 8). Among the other variants, the Mahalanobis distance contributes to the largest improvement, around 10%. Overall, CardioID improves the BAC performance by 12% with respect to the benchmark [28].

Mahalanobis contribution. A final point to discuss is the role of the Mahalanobis distance. Among all our variants, the Mahalanobis variant provides the biggest improvement,[11] which leads to the following question: If we simply replace the Euclidean distance with the Mahalanobis distance in [28] with the camera dataset, would the performance get comparable to CardioID? The result is shown in Figure 17, where we can see that there is an improvement across all signal variances but the gain is not as significant as when Mahalanobis runs on top of MS and MC. When all the signal variance is considered, the BAC of the modified SOTA ([28]+Mahalanobis) is below 80%, while that of CardioID is around 85%. Similar to what happened with identification, where we simply modify the SOTA to bypass MS and MC (cf. Figure 16), this result proves that all the foundational blocks of CardioID are important to obtain a maximum performance gain.

Summary. Based on the results with all the datasets, we can summarize the following takeaway lessons. Overall, morphology stabilization (Section 4) and the methods to overcome the non-linear effects of authentication (Section 6) improve performance, while morphology classification (Section 5) improves the acquisition rate (more users and less latency). For the public dataset with uncontrolled data, our methods can bring back the accuracy of identification and authentication close to the desired target of 95% BAC using a single cardiac period for testing. With the private and MIMIC-III dataset, however, even though our methods still have a better performance than the SOTA, they do not reach the desired 95% target. To ameliorate this problem, one could place stricter constraints on the types of morphologies that can be accepted and use more periods for testing, but that would increase the training overhead of the system and the delay while testing.

[10] In [28], the authors report a 95% BAC using a single sample. Their signals seem to have a variance of 1.5. If we extrapolate the performance of [28] in Figure 18(b), we will obtain the reported result.

[11] We also tried the variant MC+MultiCluster, without including the Mahalanobis distance, but the result of that variant was lower than the MC+Mahalanobis variant.

8 CONCLUSIONS

Motivated by the fact that PPG biometrics have been largely limited to controlled setups, we analyze the impact of more realistic *uncontrolled* signals. We identify three main limitations in the SOTA: the same filtering parameters are used to obtain the features for all users, a single dominant morphology is assumed, and there are important non-linear effects that have not been considered. Our solution, named CardioID, overcomes those limitations with a novel morphology stabilization and classification mechanism, and by using the Mahalanobis distances with a multi-cluster approach. Overall, considering a wide variety of cardiac cycles from three datasets, our work increases the *average* BAC from around 70% to 85%.

REFERENCES

[1] Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, and Prabhakar Raghavan. 1998. Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*. 94–105.

[2] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. 1999. OPTICS: Ordering points to identify the clustering structure. *ACM Sigmod Record* 28, 2 (1999), 49–60.

[3] Yoshua Bengio. 2011. How Large should be the DataSet for Training a Deep Auto Encoder? Retrieved from https://www.quora.com/How-large-should-be-the-data-set-for-training-a-Deep-auto-encoder

[4] Debnath Bhattacharyya, Rahul Ranjan, Farkhod Alisherov, Minkyu Choi, et al. 2009. Biometric authentication: A review. *International Journal of u-and e-Service, Science and Technology* 2, 3 (2009), 13–28.

[5] Angelo Bonissi, Ruggero Donida Labati, Luca Perico, Roberto Sassi, Fabio Scotti, and Luca Sparagino. 2013. A preliminary study on continuous authentication methods for photoplethysmographic biometrics. In *2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*. IEEE, 28–33.

[6] Stephen Butterworth et al. 1930. On the theory of filter amplifiers. *Wireless Engineer* 7, 6 (1930), 536–541.

[7] Marcus Carlsson, Peter Cain, Catarina Holmqvist, Freddy Stahlberg, Stig Lundback, and Hakan Arheden. 2004. Total heart volume variation throughout the cardiac cycle in humans. *American Journal of Physiology-Heart and Circulatory Physiology* 287, 1 (2004), H243–H250.

[8] Vikram Chandrasekaran, Ram Dantu, Srikanth Jonnada, Shanti Thiyagaraja, and Kalyan Pathapati Subbu. 2012. Cuffless differential blood pressure estimation using smartphones. *IEEE Transactions on Biomedical Engineering* 60, 4 (2012), 1080–1089.

[9] Anand Chandrasekhar, Mohammad Yavarimanesh, Keerthana Natarajan, Jin-Oh Hahn, and Ramakrishna Mukkamala. 2020. PPG sensor contact pressure should be taken into account for cuff-less blood pressure measurement. *IEEE Transactions on Biomedical Engineering* 67, 11 (2020), 3134–3140.

[10] Ahyoung Choi and Hangsik Shin. 2017. Photoplethysmography sampling frequency: Pilot assessment of how low can we go to analyze pulse rate variability with reliability? *Physiological Measurement* 38, 3 (2017), 586.

[11] Hugo Plácido Da Silva, Ana Fred, André Lourenço, and Anil K. Jain. 2013. Finger ECG signal for user authentication: Usability and performance. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 1–8.

[12] Simon Eberz, Nicola Paoletti, Marc Roeschlin, Marta Kwiatkowska, I. Martinovic, and A. Patané. 2017. Broken hearted: How to attack ECG biometrics. In *Network and Distributed System Security Symposium 2017*. Internet Society.

[13] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*. Vol. 96. 226–231.

[14] Umar Farooq, Dae-Geun Jang, Jang-Ho Park, and Seung-Hun Park. 2010. PPG delineator for real-time ubiquitous applications. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 4582–4585.

[15] Y. Y. Gu, Y. Zhang, and Y. T. Zhang. 2003. A novel biometric approach in human verification by photoplethysmographic signals. In *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, 2003. IEEE, 13–14.

[16] John E. Hall. 2015. *Guyton and Hall Textbook of Medical Physiology E-Book*. Elsevier Health Sciences.

[17] Alex Hern. 2017. Samsung Galaxy S8 iris scanner fooled by German hackers. *The Guardian* (2017).

[18] Geoffrey E. Hinton and Ruslan R. Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science* 313, 5786 (2006), 504–507.

[19] Steven A. Israel, John M. Irvine, Andrew Cheng, Mark D. Wiederhold, and Brenda K. Wiederhold. 2005. ECG to identify individuals. *Pattern Recognition* 38, 1 (2005), 133–142.

[20] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3, 1 (2016), 1–9.

[21] Nima Karimian, Damon Woodard, and Domenic Forte. 2020. ECG biometric: Spoofing and countermeasures. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 2, 3 (2020), 257–270.

[22] Walter Karlen, Chris Petersen, Jennifer Gow, J. Mark Ansermino, and Guy Dumont. 2013. Photoplethysmogram processing using an adaptive single frequency phase vocoder algorithm. In *Biomedical Engineering Systems and Technologies: 4th International Joint Conference (BIOSTEC'11)*, Springer, 31–42.

[23] Walter Karlen, Srinivas Raman, J. Mark Ansermino, and Guy A. Dumont. 2013. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering* 60, 7 (2013), 1946–1953.

[24] A. Reşit Kavsaoğlu, Kemal Polat, and M. Recep Bozkurt. 2014. A novel feature ranking algorithm for biometric recognition with PPG signals. *Computers in Biology and Medicine* 49 (2014), 1–14.

[25] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing* 3, 1 (2011), 18–31.

[26] Daniel D. Lee and H. Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755 (1999), 788–791.

[27] Feng Lin, Chen Song, Yan Zhuang, Wenyao Xu, Changzhi Li, and Kui Ren. 2017. 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*. 315–328.

[28] Jian Liu, Cong Shi, Yingying Chen, Hongbo Liu, and Marco Gruteser. 2019. CardioCam: Leveraging camera on mobile devices to verify users while their heart is pumping. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 249–261.

[29] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008).

[30] James MacQueen et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1. Oakland, CA, USA, 281–297.

[31] Prasanta Chandra Mahalanobis. 1936. On the generalized distance in statistics. In *National Institute of Science of India*.

[32] Karl Pearson. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* (1901).

[33] Kai Peng, Lixin Zheng, Xiaolong Xu, Tao Lin, and Victor C. M. Leung. 2018. Balanced iterative reducing and clustering using hierarchies with principal component analysis (PBIRCH) for intrusion detection over big data in mobile cloud environment. In *Security, Privacy, and Anonymity in Computation, Communication, and Storage: 11th International Conference and Satellite Workshops (SpaCCS'18)*. Springer, 166–177.

[34] Sam T. Roweis and Lawrence K. Saul. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 5500 (2000), 2323–2326.

[35] Sairul I. Safie, John J. Soraghan, and Lykourgos Petropoulakis. 2011. Electrocardiogram (ECG) biometric authentication using pulse active ratio (PAR). *IEEE Transactions on Information Forensics and Security* 6, 4 (2011), 1315–1322.

[36] Abhijit Sarkar, A. Lynn Abbott, and Zachary Doerzaph. 2016. Biometric authentication using photoplethysmography signals. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 1–7.

[37] Gholamhosein Sheikholeslami et al. 2000. WaveCluster: A wavelet-based clustering approach for spatial data in very large databases. *The VLDB Journal* (2000).

[38] Ali Siam et al. 2019. *Real-World PPG Dataset*. Mendeley Data. DOI: http://dx.doi.org/10.17632/yynb8t9x3d.1

[39] Yogendra Narain Singh and Phalguni Gupta. 2008. ECG to individual identification. In *2008 IEEE 2nd International Conference on Biometrics: Theory, Applications, and Systems*. IEEE, 1–8.

[40] Petros Spachos, Jiexin Gao, and Dimitrios Hatzinakos. 2011. Feasibility study of photoplethysmographic signals for biometric identification. In *2011 17th International Conference on Digital Signal Processing (DSP)*. IEEE, 1–5.

[41] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 5500 (2000), 2319–2323.

[42] Nima Karimian, Zimu Guo, Mark Tehranipoor, and Domenic Forte. 2017. Human recognition from photoplethysmography (ppg) based on non-fiducial features. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4636–4640.

[43] Edward Jay Wang, William Li, Doug Hawkins, Terry Gernsheimer, Colette Norby-Slycord, and Shwetak N. Patel. 2016. HemaApp: Noninvasive blood screening of hemoglobin using smartphone cameras. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 593–604.

[44] Weizheng Wang, Qing Wang, and Marco Zuniga. 2022. CardioID: Mitigating the effects of irregular cardiac signals for biometric identification. In *EWSN*.

[45] Gerhard Whitworth. 2018. *Why Do Athletes Have a Lower Resting Heart Rate?* Retrieved from https://www.healthline.com/health/athlete-heart-rate

[46] Umang Yadav, Sherif N. Abbas, and Dimitrios Hatzinakos. 2018. Evaluation of PPG biometrics for authentication in different states. In *ICB*. http://www.comm.utoronto.ca/Ëoebiometrics/PPG_Dataset/

[47] Jianchu Yao, Xiaodong Sun, and Yongbo Wan. 2007. A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 4576–4579.

[48] Zhidong Zhao, Lei Yang, Diandian Chen, and Yi Luo. 2013. A human ECG identification system based on ensemble empirical mode decomposition. *Sensors* 13, 5 (2013), 6832–6864.

Received 23 November 2022; revised 5 June 2023; accepted 5 September 2023