

# CardioID: Mitigating the Effects of Irregular Cardiac Signals for Biometric Identification

Weizheng Wang  
Delft University of Technology  
w.wang-14@tudelft.nl

Qing Wang  
Delft University of Technology  
qing.wang@tudelft.nl

Marco Zuniga  
Delft University of Technology  
m.a.zunigazamalloa@tudelft.nl

## Abstract

Cardiac patterns are being used to obtain hard-to-forge biometric signatures and have led to high accuracy in state-of-the-art (SoA) 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 (i.e., irregularity) makes it harder to obtain stable and distinct user features. Furthermore, SoA usually fails to identify specific groups of users, rendering existing identification methods futile in uncontrolled scenarios. To solve these problems, we propose a framework with three novel properties. First, we design an adaptive method that achieves stable and distinct features by tailoring the filtering spectrum to each user. Second, we show that users can have multiple cardiac morphologies, offering us a much bigger pool of cardiac signals and *users* compared to SoA. Third, we overcome other distortion effects present in authentication applications with a multi-cluster approach and the Mahalanobis distance. Our evaluation shows that the average balanced accuracy (BAC) of SoA drops from above 90% in controlled scenarios to 75% in uncontrolled ones, while our method maintains an average BAC above 90% in uncontrolled scenarios.

## 1 Introduction

Biometrics play a fundamental role in human identification. The popular systems rely on external features, such as fingerprints, iris and audio patterns, and face contours. These systems have excellent precision but they are vulnerable to attacks: fingerprints can be recreated in latex from touched objects [15]; iris patterns can be scanned and emulated [11]; audio patterns can be spoofed based on the daily vocal segments [2]; pictures from the Internet can be used to obtain renditions that can fool face recognition systems [3].

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. 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 [14]; another option is to place a finger on top of the flashlight and camera in a smartphone [18]. 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 [9, 15, 18, 25].

**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 (SoA) approaches. This uncontrolled environment is common in our targeted scenarios: smartphone and door login systems.

**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 them. Our work has four main contributions:

*Contribution 1: Morphology Stabilization [Section 3].* The biometric information present in cardiac cycles is restricted to a narrow spectrum of the signal. A key limitation of the SoA 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 filtering 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 4].* SoA 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 SoA assumption of considering a single dominant morphology leaves out users that rarely have such cardiac morphology, rendering SoA methods futile for those scenarios. We show that a *single user* can have *multiple* valid morphologies. Our ability to consider a wider range of morphologies reduces the amount of time required to test a system, increases the 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 5].* The SoA utilizes PPG signals to perform two types of biometric applications: identification and authentication. For identification, SoA uses linear (PCA [18], LDA [25,31]) and non-linear approaches (NN-based [14, 28]), but there is no analysis determining what approach is better and why. We show that if we tackle the non-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 SoA methods: the use of Euclidean distances and the assumption that the features of a subject form a *single* cluster. We propose to leverage *multi-cluster* and Mahalanobis distance to ameliorate non-linear effects.

*Contribution 4: Thorough multi-sensor and multi-application evaluations [Section 6].* 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 (usually identification with pulse oximeters), *no study has evaluated all four*. Our evaluation includes both types of sensors and applications, and considers cardiac cycles over a wide variability range, from low-variance (controlled) to high-variance (uncontrolled). Overall, we show that the SoA performs well in controlled scenarios, the average balanced accuracy (BAC) for identification and authentication is above 90%; but in uncontrolled scenarios, it drops to 75%. Our methods recoup the average BAC above 90% in uncontrolled scenarios.

## 2 Preliminaries

### 2.1 PPG basics

The cardiac cycle represents the continuous 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 [10], and different tissue thickness and blood vessel distribution [10], the cardiac signal has been used to obtain unique biometric signatures [13]. 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.) [15], or the spectral information in the frequency domain [14], are optional features used to perform identification.

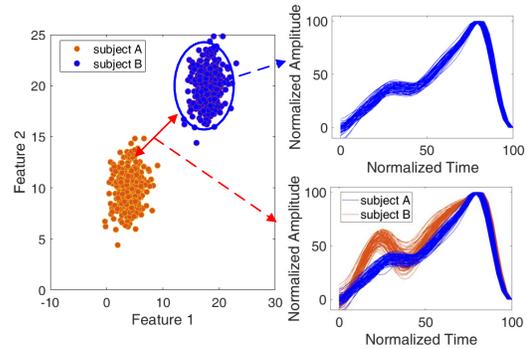


Figure 1: Sample application with *controlled* PPG signals.

### 2.2 Applications, morphologies and metrics

We analyze the performance of two different 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 SoA has tested with both applications: they only focus on one, usually on identification. Our study analyzes both. Independently of the target application, achieving high biometric accuracy with PPG signals requires achieving 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 two users’ PPG signals collected in a controlled manner. Under these favorable cases, it is simple to tackle the above challenges and differentiate the individuals.

**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 SoA to measure accuracy. Some studies use the equal error rate (EER) [14], others use F1-score [15] or BAC [18]. All these metrics are derived from true/false positive/negative results. Our results can be presented in any of these metrics. We decide to use BAC because our datasets are imbalance. BAC is the average of the true positive rate (TPR, sensitivity) and the true negative rate (TNR, specificity):  $BAC = (TPR + TNR)/2$ , where  $TPR = \text{True Positive}/(\text{True Positive} + \text{False Positive})$ , and  $TNR = \text{True Negative}/(\text{True Negative} + \text{False Positive})$ .

### 2.3 The detrimental effect of irregular cycles

Multiple PPG studies report a high identification accuracy, ranging from 85% to almost 100%, depending on vari-

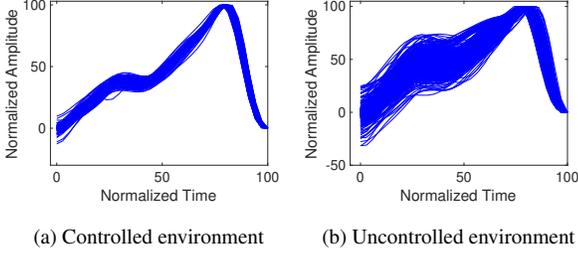


Figure 2: Cardiac periods collected for the same subject.

Table 1: Performance of SoA.

	Controlled	Uncontrolled
Signal Variance	1.83	2.96
BAC for identification [15]	91%	72%
BAC for authentication [18]	93%	69%

ous evaluation parameters and scenarios [9, 14, 15, 18, 25, 28]. 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 dataset situation, and 2) the individuals in the dataset. The controlled datasets typically focus on healthy individuals with a narrow age range between 20 and 40 (cf. Subsection 6.1). Thanks to their good cardiac status, their cardiac signals are prone to be stable. For each individual in the dataset, they take the measurement without minor hand movements and the unconscious pressure change between the fingertip and sensor. So their cardiac signals are even more stable.

In contrast to the *controlled* process, the uncontrolled process is a more realistic daily usage scenario. It covers the age from children to elders and includes minor hand movements and unconscious pressure. Figure 2 depicts PPG signals for a single individual collected in controlled and uncontrolled environments. The small variance observed in Figure 2a is similar to that in Figure 1 of [25] and Figure 3 of [18].<sup>1</sup> While it is reasonable 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 SoA methods. Table 1 shows a preliminary evaluation with four subjects, for whom we collected PPG signals in controlled and uncontrolled environments. The extraction processes comply with the corresponding papers. In identification evaluation [15], there are 758 and 1822 periods for the controlled and uncontrolled environment respectively. In authentication evaluation [18], there are 640 and 1631 periods for the controlled and uncontrolled environment. The exact description of the SoA methods and the means used to calculate the signal variance are explained in Subsection 6.2 Subsection 6.1, respectively. *The important takeaway is that when the SoA is tested with controlled data, the performance is high, as reported in the original studies; but when tested with highly variable signals, the accuracy drops significantly.*

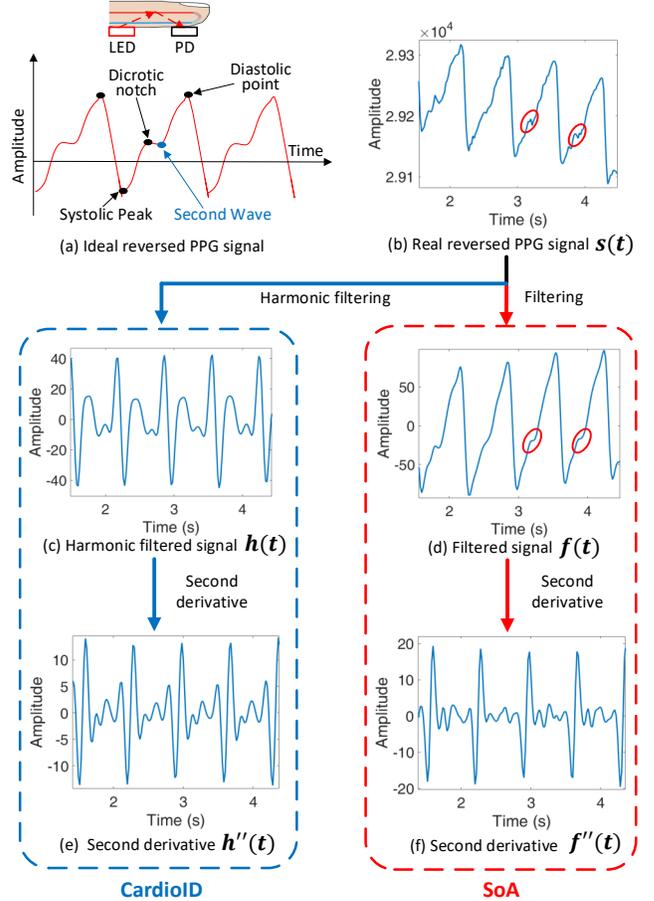


Figure 3: Morphology stabilization.

### 3 Morphology Stabilization

A major shortcoming of the SoA 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 SoA. First, we describe the methods we borrow from the SoA (Subsection 3.1), and then, we describe their limitation and present our contributions (Subsection 3.2).

#### 3.1 From SoA: Basic filtering and derivatives

Figure 3a depicts an ideal PPG signal. The biometric signature of an individual is captured by four fiducial points: diastolic (highest valley), systolic (lowest peak), the dicrotic notch (which form a small peak in the middle of the period) and second wave. Figure 3b 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 can be so high 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

<sup>1</sup>These studies do not post their PPG data. To infer the variance of their signals, we have to rely on their figures.

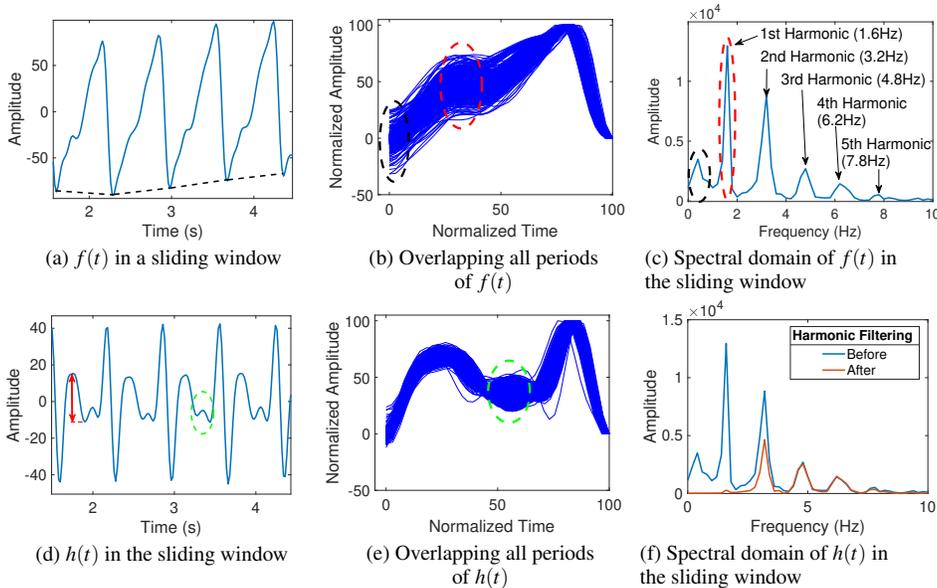


Figure 4: Frequency analysis to determine the lower cut-off frequency  $f_l$ .

points limits the number of features. To overcome these effects, the SoA proposes a basic filtering step and the use of the second derivative of the PPG signal.

**Filtering.** To mitigate the noise in PPG signals, SoA 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 [30], the lower cut-off frequency  $f_l$  is usually set to that value. Regarding the upper cut-off frequency  $f_h$ , according to [6], sampling frequencies above 25 Hz do not provide any extra information, hence,  $f_h$  can be set to 12.5 Hz (due to Nyquist sampling theorem). Some studies use other filtering bands [15, 18], but the overall filtering process is similar. Figure 3d shows a PPG signal  $f(t)$  after being filtered with a second-order Butterworth bandpass filter with bandwidth 0.5-12.5 Hz [4].

**Derivatives.** Filtering alleviates noise, but it also eliminates valuable information. For instance, the raw PPG signal  $s(t)$  in Figure 3b contains faint but detectable second waves (red circles). After filtering, however, those fiducial points no longer exist (red circles in Figure 3d). To overcome this issue, researchers obtain features not only from  $f(t)$  but also from its second derivative  $f''(t)$  [15]. Figure 3f depicts the  $f''(t)$ , which exhibits more fiducial points than  $f(t)$ .

### 3.2 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 user. This process allows us to obtain more stable morphologies for every user (Challenge 1) and distinct fiducial points among users (Challenge 2). Considering our harmonic filtering depends on the subjects' heart rates that can change over time, our system proceeds signals based on a 5-second sliding window with a 1-second stride, to track the change of subjects' heart rates.

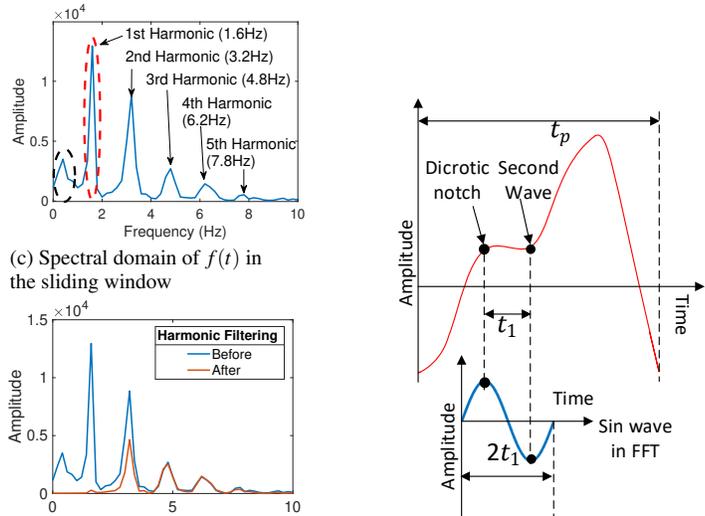


Figure 5: Frequency analysis to determine the upper cut-off frequency  $f_h$ .

#### 3.2.1 Determining the lower cut-off frequency $f_l$

The SoA 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 4a shows the filtered signal  $f(t)$  using SoA methods and Figure 4b shows the overlapping cardiac cycles using the end point of periods as an alignment anchor. We can observe a large variance in the starting points (black ellipsoid in Figure 4b) and significant instability in the dicrotic notch (red ellipsoid in Figure 4b).

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 4c. 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 4a), which causes the height differences between the starting and end points in periods. Considering the end points 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 is concentrated around the harmonics of the heartbeat, in particular the first harmonic (red ellipsoid in Figure 4c). The SoA does not filter the first harmonic for its 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. Furthermore, it is unnecessary to keep the first harmonic to obtain the heart rate because it is still contained on the other harmonics.

Our analysis indicates that to ameliorate 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  $f_l = 2f_{1h}$ .

### 3.2.2 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 4c 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 heartrate, 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 4c prove the correctness of our analysis.

### 3.2.3 Adaptive filtering

The frequency response of our filter is solely based on 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 heartrate instead of fixed parameters, allowing us to perform accurate adaptive filtering per subject. Figure 4d-f show the signals filtered with our method, their overlapping cycles and spectral domains. We can observe that, compared to the filtered signal  $f(t)$  in SoA,  $h(t)$  has three advantages: (i) the signal variance is much lower throughout the entire cycle, Figure 4e; (ii) the difference between the second wave and dicrotic notch is accentuated significantly, red arrow in Figure 4d; and (iii) our method exposes another fiducial point, green ellipsoid in Figure 4d, which we can exploit to obtain more features as explained next.

### 3.2.4 Derivatives

As described earlier, the SoA 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  $f''(t)$ . 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

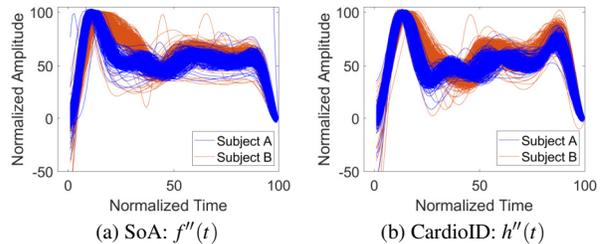


Figure 6: Overlapping periods with uncontrolled data.

Figure 6b shows that the second derivative disentangles the ‘knot’ caused by the new fiducial point captured by the green ellipsoid in Figure 4e.

**Summary.** Overall, our approach also follows the two basic steps of the SoA, 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. SoA obtains its features from  $f(t)$  and  $f''(t)$ , and we obtain them from  $h(t)$  and  $h''(t)$ . An exact description of what features we use is provided in the next section.

## 4 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 show that cardiac periods can have multiple morphologies, then we describe the features derived from multiple morphologies.

### 4.1 Multiple morphologies

Currently, all studies using fiducial points assume a single macro morphology for all subjects. That is valid in controlled scenarios, but in uncontrolled scenarios various factors can cause the appearance of multiple morphologies: unintended fingertip pressure [5], significant differences in the cardiac profiles of subjects, etc. When we perform the second derivative of our harmonic signal  $h(t)$ , we observe multiple morphologies. Figure 7 depicts the three most dominant macro morphologies observed in  $h''(t)$ :  $h''_1(t)$ ,  $h''_2(t)$ ,  $h''_3(t)$ . Those dominant morphologies account for 98.4% of the periods measured in a *public dataset* with 35 subjects [27], 15301 out of 15557 periods and 97.5% of the periods measured in a *private dataset* with 43 subjects, 11328 out of 11617 periods. Figure 8 shows the relative presence of those three macro morphologies in the datasets. There are other macro morphologies in  $h''(t)$ , but we do not consider them because they are rarely present (“Discard” label).

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*, the system latency of identifying a subject is long, because the system will need to wait for the right morphology to arrive. For example, for the

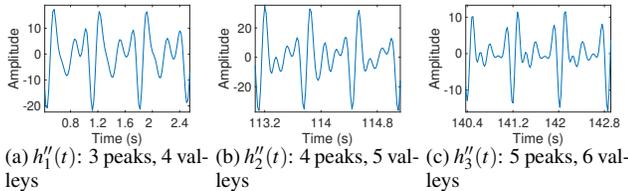


Figure 7: Dominant morphologies for  $h''(t)$ .

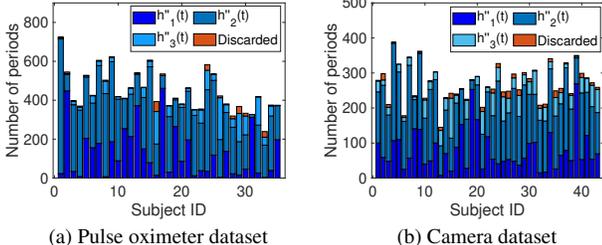


Figure 8: Frequency of occurrence of morphologies.

public dataset, our system can obtain 1.229 (15301/12444) 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 one of the subject(s) 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.

## 4.2 Feature extraction

We extract features from  $h(t)$  and  $h''(t)$ . Like several other studies in the area [15, 18], our features are largely based on the geometric relations amongst fiducial points. Figure 9 and Table 2 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 9a, 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 features (differences between contiguous fiducial points). Figure 9b 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, respectively. Features based on duration and height are susceptible to heartbeat variance. In [13, 18], the authors state that normalizing the features makes them immune to heart rate changes. Therefore, we also normalized the duration and height features of  $h(t)$  and  $h''(t)$ .

It is important to highlight that 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 val-

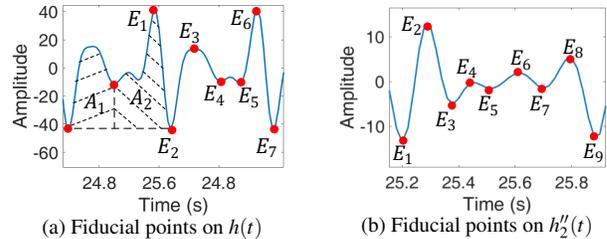


Figure 9: Fiducial points used to extract features.

Table 2: 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	$(-1)^{i+1} * H''_i/D''_i$

ley. However, contrary to the multi-morphology approach used for  $h''(t)$  in Subsection 4.1, 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 accuracy of the system. 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 2.

## 5 Identification and Authentication

As stated earlier, SoA 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.

### 5.1 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 SoA utilizes two kinds of 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) [31], which simultaneously reduces dimensionality and draws decision boundaries. The most representative non-linear methods are based on neural networks (NN) [14]. As it is customary with NN [12], we first use an autoencoder for dimensionality reduction, blue layers in Figure 10, 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

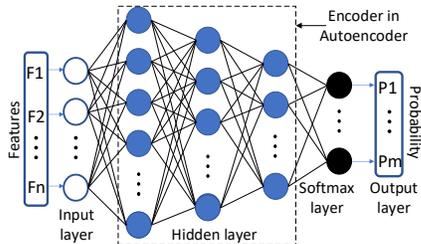


Figure 10: The structure of our neural network.

morphology (each pipeline receives the corresponding set of features presented in Table 2). 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 network structure and parameter values, we tailor three different NN modules for each morphology. The hidden (blue) layer neurons for morphologies one, two and three are 128-64-32, 170-85-42 and 128-64-32, respectively. The activation functions of neurons are sigmoid to guarantee the system’s non-linearity, and parameters such as L2 and sparsity regularization are tuned for each morphology.

## 5.2 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 (to identify informative features and reduce the number of features in case of curse of dimensionality) 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 SoA for dimensionality reduction, and then, some techniques to improve the definition of boundaries.

**Dimensionality reduction.** It can also be performed with linear and non-linear methods. There are two mainstream linear techniques: principal component analysis (PCA) [22] and non-negative matrix factorization (NMF) [17]. NMF requires non-negative features, but the slopes in our feature set can be negative. Hence, similar to prior studies [18], we also adopt PCA. Even though there are several non-linear dimensionality reduction techniques –such as Isomap [29], local linear embedding (LLE) [24], t-distributed stochastic neighbor embedding (t-SNE) [19], and autoencoder [12]– we did not find SoA 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. 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 counts with several thousand samples coming from *all users*, but in authentication, we only have a few hundred samples coming from a *legitimate user*. Due to this finding, in our evaluation, we only consider PCA for authentication.

**Mahalanobis distance.** After dimensionality reduction, the most significant features of a subject usually form a clus-

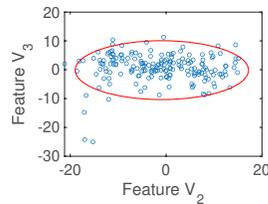


Figure 11: Mahalanobis.

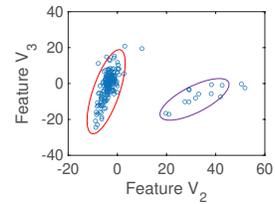


Figure 12: Multi-cluster.

ter similar to the one shown in Figure 11. 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 [18]. But Euclidean distances are fundamentally ill-equipped to deal with feature spaces that have widely different variances. For example, in Figure 11, 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  $v_3$ , causing numerous false positives; or too short for  $v_2$ , causing significant false negatives. Therefore we adopt the well-known Mahalanobis distance [21], which considers the standard deviations in each dimension and can be used to define tight boundaries such as the red ellipsoid shown in Figure 11.

**Multi-cluster approach.** Current PPG authentication systems assume that the features of a user converge to a *single* cluster [18, 31]. However, with uncontrolled data, we observed that a single subject can form two or more clusters for a single morphology, as depicted in Figure 12. We need an authentication system that can identify multiple clusters and then use the Mahalanobis distance to set an appropriate threshold for each cluster<sup>2</sup>.

For our purposes, the cluster method should meet three requirements: (i) be resilient to the presence of outliers, (ii) able to detect clusters with arbitrary shape, and (iii) fast. Hierarchical cluster methods, such as BIRCH [23]; and centroid-based methods like K-means [20] are vulnerable to outliers and cannot detect arbitrary shapes. Most grid-based cluster methods, like CLIQUE [1], and density-based methods, like DBSCAN [7] do not have shortcomings (i) and (ii), but they need a relatively long computation time. Due to the above reasons, we decide to use WaveCluster [26], 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  $O(n)$ .

## 6 Performance Evaluation

### 6.1 Datasets

We use two datasets to evaluate the performance of CardioID. The first dataset, using *pulse oximeter* for signal collection, is public and recently published in [27]. All subjects are sitting during the signal collection. The second dataset is collected by us.<sup>3</sup> We use an iPhone 7 camera to gather cardiac periods from sitting volunteers. The camera records videos of the redness change on our volunteers’ fingertips attached to the camera at 60 FPS. In each frame, we focus

<sup>2</sup>To calculate Mahalanobis distance, the number of samples must be greater than the number of features. If a cluster has few samples like the purple one in Figure 12, we use spline interpolation to add extra points.

<sup>3</sup>This data collection is approved by the HERC of our university.

Table 3: Details of the two 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

only on the red channel of the pixels covered by the fingertip. The method to select the covered pixels is the same as in [18]:  $I_{red} > 80\% \times (I_{red} + I_{blue} + I_{green})$ . Then, 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 as auto.

**Significance of datasets.** The parameters of the three datasets –the number of subjects, gender, age distribution (average, standard deviation and range), the recording duration (RD) and variability (cross-track error, CTE)– are given in Table 3. There are two important points to highlight about the selection of our datasets.

First, *no SoA 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 the noisy motion artifacts [8]. This is one reason why the CTE (variability) in Table 3 is higher for the camera dataset.

Second, *our first two datasets consider a broader group of people, which allows us to consider a wide range of dynamics and prevent the system from overfitting a narrow segment of users*. Even for *healthy* people, which is the main focus of our work and most of the SoA, the morphology of cardiac signals can vary significantly based on the age group and skin tone. This is another reason why the CTE in Table 3 is higher for the camera dataset. Our baseline studies, [18] and [15], focus on a narrow age segment of the adult population, 22-33 and 18-46, respectively. The results obtained with these concentrated age distributions are hard to validate for people of all ages. The age ranges of our datasets 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 vs. 25, [18]) and 17% more for the pulse oximeter (35 vs. 30, [15]). 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.

**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 13; and then, we calculate the cross-track error (CTE)<sup>4</sup> from every (blue) PPG signal to its average. Denoting  $e_i$  as the CTE for signal  $i$ , the signal variance for a

<sup>4</sup>In GPS system, CTE measures the distance from the current position to the given path. We tried different similarity metrics, including dynamic time warping, and we found that CTE renders the best in PPG signals.

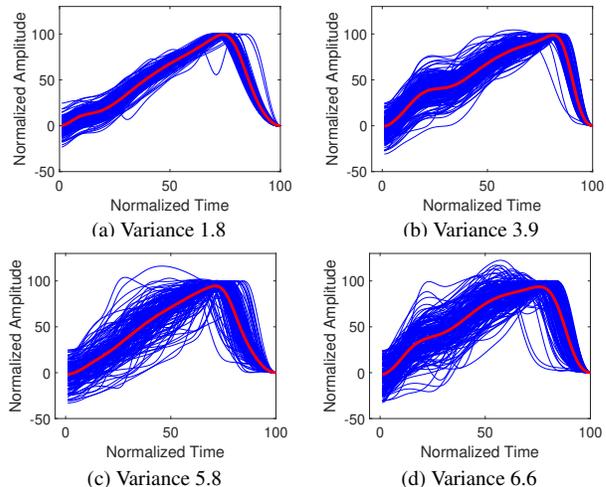


Figure 13: Different signal variances.

subject is the mean absolute error for all  $e_i$ ’s. The average signal variances for our datasets are 4.29 and 5.97, shown in Table 3. *In order to put these values in context, it is important to note that the variance found in SoA plots is a bit lower than what is shown in Figure 13a, less than 2*. The majority of users in the public dataset have a variance in the range [2, 6] and in private dataset in [4, 6]. Hence, our evaluation copes with a spectrum of signal variability that has not been tackled before.

## 6.2 Baselines used for comparison

We utilize two SoA studies as baselines for comparison, one for identification [15] and the other for authentication [18]. The reasons for selecting those baselines are presented in Section 7. In this subsection, we quantify the acquisition rate improvement of CardioID and the performance of our work and the SoA baselines.

### 6.2.1 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 4, 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  $S$  are equal to 14347 and 10728 periods for the public and private datasets, respectively, we first need to find  $S'$  for the SoA baselines. The morphology and features used by Kavsaoglu *et. al.* are presented in Figures 7 and 8 in [15], and the corresponding information for Cardio-Cam is provided in Figure 8 in [18].<sup>5</sup> We use that information to discard the morphologies that do not conform to their

<sup>5</sup>Upon close inspection, we notice that, in both SoA studies, the second derivative of  $f(t)$  is a signal similar, but not exactly the same, as morphology-2 in our case,  $h_2'(t)$ .

requirements (the right morphology is necessary to obtain their features). After discarding the non-conforming morphologies in SoA, we obtain the following acquisition rates: 74.6% (public dataset) and 64.5% (private dataset) for [15]; and 59.2% (public) and 32.8% (private) for [18]; significantly lower than the 98.4% (pulse oximeter) and 97.5% (camera) 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 [18], and thus, there is no possibility to authenticate them with that method.*

Moreover, the solutions in these SoA [15, 18] have a long acquisition delay. There are 12444, 10320, and 3000 seconds for the public dataset, private dataset and MIMIC-III dataset, respectively. The acquisition speeds are 0.932 (public dataset) and 0.726 (private dataset) periods/second for [15]; and 0.740 (public) and 0.369 (private) periods/second for [18]; significantly lower than the 1.229 (pulse oximeter) and 1.097 (camera) periods/second obtained for CardioID. Those SoA acquisition speeds are below 1 period/second. Some of the acquisition speeds are even lower than 0.5 period/second, which is unfriendly for users.

### 6.2.2 CardioID variants

To assess the impact of our contributions –*morphology stabilization, morphology classification and the reduction of non-linear effects for authentication*– we create different variants for CardioID. SoA approaches are implemented based on 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 2. This variant only considers morphology-2 periods,  $h_2''(t)$ . The classification method is K-NN, the same as in [15].
- 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 [18].
- 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.

### 6.2.3 Emulating a wide range of signal variances

Our aim is to evaluate the SoA 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 de-

Table 4: Percentage of detectable subjects and periods in the public (pulse-oximeter) dataset.

Public	Signal variance	2	4	6	All
[15]	Subject %	82.9	100	100	100
	Period %	21.6	54.2	67.3	74.6
[18]	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

Table 5: Percentage of detectable subjects and periods in the public (camera) dataset.

Private	Signal variance	2	4	6	All
[15]	Subject %	37.2	90.7	100	100
	Period %	5.5	33.1	51.2	64.5
[18]	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

cide to divide 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 Table 4 and Table 5. For example, if we look at Table 4 for reference [15], 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 [15]; and (ii) only 29 subjects, out of 35 (82.9%), have more than 20 periods satisfying the conditions in (i).

Table 4 and Table 5 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 SoA 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 [18] cannot include all users in both datasets. This is an important point showing that the requirements in [18] for the single morphology is 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 SoA).* The broad type of users in our dataset enables us to expose this age limitation. The variants, LDA, NN, Mahal and Cluster, have the same percentage of subjects and periods as MC because they are derived from that variant.

### 6.2.4 Providing the right context for evaluation

Until now, we have only evaluated the impact of CardioID on the acquisition rate. In the remainder of this section we evaluate its performance, and it is important to consider the following points: (i) for each emulated subset, we use the

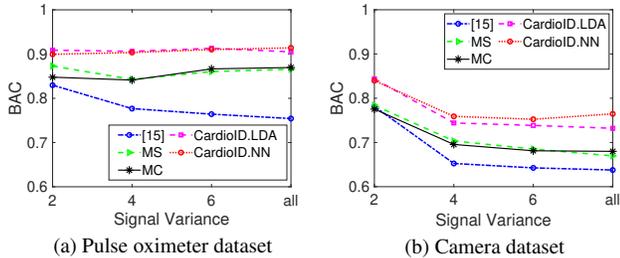


Figure 14: Identification performance.

first 80% of periods for training and the rest for testing; (ii) the variance observed in SoA 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 SoA 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 SoA reports a BAC of 0.95 [18], 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 approach 0.95 BAC with *uncontrolled signals*. In authentication and identification, *improvements in the order of 5%, or above, are already considered highly significant* [3].

### 6.3 Identification

**Public dataset.** Figure 14a shows the results for the pulse oximeter dataset. The MS variant provides the most significant improvement, above 10% for most signal variances. 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 the fact that it accepts multiple morphologies improves inclusion (more subjects are accepted) and the acquisition rate (40% higher than the MS variant and 20% higher than the SoA baseline for the entire dataset [15], cf. last column in Table 4). 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 SoA drops to 76.1% [15], leading to a total improvement of around 15%.

**Private dataset.** Figure 14b 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 5), and the classification methods (LDA and NN) provide around a 7% improvement.

It is important to note that even though CardioID performs significantly better than the SoA 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.

**Linear vs. non-linear methods.** SoA studies have been using linear [25, 31] and non-linear [14, 28] 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 14). 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 [15] with LDA in the camera dataset (while maintaining everything else the same). The results in Figure 15 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.

### 6.4 Authentication

**Public dataset.** Before discussing the authentication results, we need to highlight a critical difference compared to identification: for authentication, the SoA [18] fails to include all types of users. In Table 4, we can observe that both, the baseline for identification [15] 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 [18] (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 [18] are so stringent that we cannot get enough periods for 6 subjects. Hence, in authentication, CardioID faces a more challenging scenario than the SoA because a bigger population increases the likelihood of errors.

With that clarification, we can now discuss the main insights for the pulse oximeter dataset in Figure 16a. 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, however, we can observe that 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 [18] achieves 11% less BAC with 29 subjects.

**Private dataset.** Figure 16b shows the results with the camera dataset. Due to the higher signal variability of this dataset, the SoA baseline filters even more periods than with the prior dataset: the percentage of subjects for the SoA is less in Table 5 than in Table 4 (69.8% vs. 82.9%), and as stated before, three subjects did not have a single period sat-

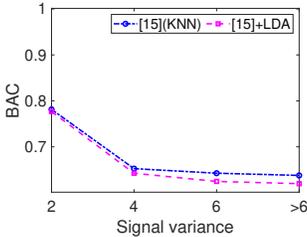
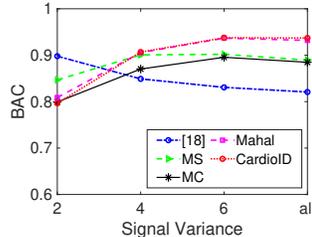
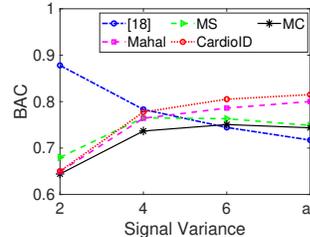


Figure 15: LDA contribution.



(a) Pulse oximeter dataset.



(b) Camera dataset.

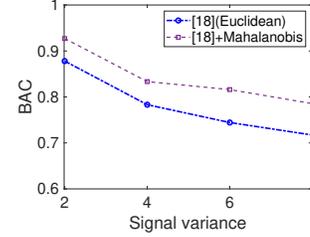


Figure 17: Mahal. contribution.

isfying the morphological requirements in [18], which would make the SoA system futile for this target group. Even if we leave that critical point aside, CardioID still outperforms the SoA 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.

There is one result, however, that we did not expect and it highlights a particular strength of the SoA with the camera: for  $t = 2$ , [18] has a strong performance compared to all of our variants, and it is in accordance with what the authors report in the original paper.<sup>6</sup> Initially, we thought that it was because, at  $t = 2$ , the SoA considers only 11 out of 43 subjects (25.6% in Table 5), but our MS variant, which relies only on morphology-2, also considers the same amount of subjects and performs worse. The SoA has stringent morphological requirements that filter out too many cardiac periods when the signal variability is high, but this conservative standard allows them to have more similarity among their periods when the data is more controlled, and it 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.

**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,<sup>7</sup> which leads to the following question: If we simply replace the Euclidean distance with the Mahalanobis distance in [18] in 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 variances but the gain is not as significant as when Mahalanobis runs on top of MS and MC. When all signals are considered, the BAC of modified SoA ([18]+Mahalanobis) is below 80%, while that of CardioID is around 85%. Similar to what happened with identification, where we simply modify the SoA to bypass MS and MC (cf. Figure 15), 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 datasets, we can

<sup>6</sup>In [18], 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 [18] in Figure 16b, we will obtain the reported result.

<sup>7</sup>We also tried the variant MC+MultiCluster without including the Mahalanobis distance, but the result is lower than the MC+Mahalanobis variant.

summarize the following takeaway lessons. Overall, morphology stabilization (Section 3) and the methods to overcome the non-linear effects of authentication (Section 5) improve performance, while morphology classification (Section 4) 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 SoA, 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.

## 7 Related Work

We divide the related work into three main phases, highlighting the elements we build upon from the SoA. A summary of the most relevant studies is presented in Table 6.

*Phase 1: Basic identification.* Gu *et al.* report the first results for PPG identification using just four features [9]. They achieve an accuracy of 94% using a discriminant function. Later, researchers found that the derivatives of a PPG signal can provide more stable and unique features [32]. 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.* [28] considers fiducial features and derivatives with two data sets. They report widely different performance for each set, EER (Equal Error Rate) 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*”. One of the most comprehensive evaluations is performed in [15]. They use 40 features from PPG signals and their derivatives (first and second), and utilize a *single period* for testing to obtain an accuracy of 87.2% F1 score, which is stricter than 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 baseline of comparison for identification.

*Phase 2: Non-fiducial features and more challenging PPG signals.* An important motivation for our study comes from [25] and [31]. Sarkar *et al.* [25] 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 test-

Table 6: Most relevant SoA studies. If a public dataset is also used in some studies, a reference is given in ‘# subjects’ column. The studies in bold are used as comparison baselines. FI stands for fiducial features; NFI stands for non-fiducial features.

	Application	Sensor Type	# Subjects	# Features	Decision Method	# PT	Accuracy
<b>[15], 2014</b>	Ident.	pulse oximeter	30	40 (FI)	KNN	1	87.0% (F1)
[25], 2016	Ident.	pulse oximeter	32 [16]	15 (FI)	LDA & QDA	5	92.5% (Rank-1)
<b>[18], 2019</b>	Auth.	camera	25	30 (FI) & 36 (NFI)	PCA	1	95.8% (BAC)
CardioID	Both	Both	35 [27]	32,38,44 (FI)	LDA & NN (Ident.)	1	93.2% (Ident.), 95.5% (Auth.) – (BAC)
			43		PCA (Auth.)		78.2% (Ident.), 85.3% (Auth.) – (BAC)

ing sample (~15 sec). Yadav *et al.* [31] 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. It is more resilient to noise than fiducial features, but they still require long testing sequences, between 8 and 40 cardiac periods (~6 to 60 sec). 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 non-fiducial (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 has been able to use smartphone *cameras* to attain *authentication* [18] (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. We implement the *signal processing chain* of CardioCam (filters, features and PCA method) and show that its performance decreases significantly with irregular PPG signals. Furthermore, for some users, the requirements of the CardioCam’s morphology are so strict that they cannot use the system.

## 8 Conclusions

Motivated by the fact that the study of PPG biometrics has been largely limited to controlled setups, we analyze the impact of more realistic *uncontrolled* signals. We identify three limitations in the SoA: 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 are not 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. On average, with uncontrolled signals, CardioID provides a BAC of 90%, while the SoA attains a much lower BAC of 75%.

## 9 References

- [1] R. Agrawal *et al.* Automatic subspace clustering of high dimensional data for data mining applications. In *SIGMOD*, 1998.
- [2] T. Armerding. Vocal theft on the horizon, 2017. [www.csoonline.com/article/3196820/vocal-theft-on-the-horizon.html](http://www.csoonline.com/article/3196820/vocal-theft-on-the-horizon.html).
- [3] D. Bhattacharyya *et al.* Biometric authentication: A review. *IJUNESST*, 2009.
- [4] S. Butterworth *et al.* On the theory of filter amplifiers. *Wireless Engineer*, 1930.
- [5] A. Chandrasekhar *et al.* Ppg sensor contact pressure should be taken into account for cuff-less blood pressure measurement. *TBME*, 2020.
- [6] A. Choi and H. Shin. Photoplethysmography sampling frequency: pilot assessment of how low can we go to analyze pulse rate variability with reliability? *Physiological measurement*, 2017.
- [7] M. Ester *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, 1996.
- [8] U. Farooq *et al.* Ppg delineator for real-time ubiquitous applications. In *EMBC*, 2010.
- [9] Y. Y. Gu, Y. Zhang, and Y. T. Zhang. A novel biometric approach in human verification by photoplethysmographic signals. In *ITAB*, 2003.
- [10] J. E. Hall. *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2015.
- [11] A. Hern. Samsung galaxy s8 iris scanner fooled by german hackers. *The Guardian*, 23, 2017.
- [12] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.
- [13] S. Israel *et al.* Ecg to identify individuals. *Pattern recognition*, 2005.
- [14] N. Karimian *et al.* Human recognition from photoplethysmography (ppg) based on non-fiducial features. In *ICASSP*, 2017.
- [15] A. R. Kavsaoglu *et al.* A novel feature ranking algorithm for biometric recognition with ppg signals. *Computers in biology & medicine*, 2014.
- [16] S. Koelstra *et al.* Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 2011.
- [17] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [18] J. Liu *et al.* Cardiocam: Leveraging camera on mobile devices to verify users while their heart is pumping. In *MobiSys*, 2019.
- [19] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- [20] J. MacQueen *et al.* Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967.
- [21] P. C. Mahalanobis. On the generalized distance in statistics. In *National Institute of Science of India*, 1936.
- [22] K. Pearson. 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.
- [23] K. Peng *et al.* Balanced iterative reducing and clustering using hierarchies with principal component analysis (pbirch) for intrusion detection over big data in mobile cloud environment. In *SpaCCS*, 2018.
- [24] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.
- [25] A. Sarkar *et al.* Biometric authentication using photoplethysmography signals. In *BTAS*, 2016.
- [26] G. Sheikholeslami *et al.* Wavecluster: a wavelet-based clustering approach for spatial data in very large databases. *VLDB Journal*, 2000.
- [27] A. Siam *et al.* Real-world ppg dataset. Mendeley data, 2019. <http://dx.doi.org/10.17632/yynb8t9x3d.1>.
- [28] P. Spachos *et al.* Feasibility study of photoplethysmographic signals for biometric identification. In *ICDSP*, 2011.
- [29] J. B. Tenenbaum *et al.* A global geometric framework for nonlinear dimensionality reduction. *science*, 2000.
- [30] G. Whitworth. Why do athletes have a lower resting heart rate?, 2018. <https://www.healthline.com/health/athlete-heart-rate>.
- [31] U. Yadav *et al.* Evaluation of ppg biometrics for authentication in different states. In *ICB*, 2018.
- [32] J. Yao *et al.* A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier. In *EMBC*, 2007.