Adaptive Biometric Systems: Review and Perspectives

Paulo Henrique Pisani, Abir Mhenni, Romain Giot, Estelle Cherrier, Norman Poh, André C.P.L.F. de Carvalho, Christophe Rosenberger, Najoua Essoukri Ben Amara

Adaptive Biometric Systems: Review and Perspectives

PAULO HENRIQUE PISANI[*†], Universidade Federal do ABC (UFABC), Av. dos Estados, 5001, Santo André, Brazil and Universidade de São Paulo (USP) - Instituto de Ciências Matemáticas e de Computação, Av. Trabalhador São Carlense, 400, São Carlos, Brazil
ABIR MHENNI[‡], Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse, LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisie
University of Tunis El Manar, ENIT, BP 94, Rommana 1068 Tunis, Tunisia
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
ROMAIN GIOT[§], Univ. Bordeaux, LaBRI, CNRS, UMR 5800, F-33400 Talence, France
ESTELLE CHERRIER, Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
NORMAN POH, Truststamp, Atlanta, Georgia 30308, USA
ANDRÉ CARLOS PONCE DE LEON FERREIRA DE CARVALHO[¶], Universidade de São Paulo (USP) - Instituto de Ciências Matemáticas e de Computação, Av. Trabalhador São Carlense, 400, São Carlos, Brazil
CHRISTOPHE ROSENBERGER, Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
NAJOUA ESSOUKRI BEN AMARA, Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse, LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisie;

With the widespread of computing and mobile devices, authentication using biometrics has received greater attention. Although biometric systems usually provide good solutions, the recognition performance tends to be affected over time due to changing conditions and aging of biometric data, that results in intra-class variability. Adaptive biometric systems, which adapt the biometric reference over time, have been proposed to deal with such intra-class variability. This paper provides the most up-to-date and complete discussion on adaptive biometrics systems we are aware of, including formalization, terminology, sources or variations that motivates the use of adaptation, adaptation strategies, evaluation methodology and open challenges. This field of research is sometimes referred to as template update.

CCS Concepts: • **Security and privacy → Biometrics**; • **Computing methodologies → Supervised learning**; **Semi-supervised learning settings**;

Additional Key Words and Phrases: Adaptive biometric systems, Template update, Biometric reference adaptation, Evaluation methodology, Template aging.

[*]Current address: Universidade Federal do ABC (UFABC), Santo André, Brazil
[†]This work was partially supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - grant 2012/25032-0).
[‡]This work was partially supported by Ministry of Higher Education and Scientific Research of Tunisia
[§]This work was partially supported by funds from LaBRI for emerging projects.
[¶]This work was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - grant 2013/07375-0).

DOI: 10.1145/nnnnnnn.nnnnnnn

ACM Reference format:
Paulo Henrique Pisani, Abir Mhenni, Romain Giot, Estelle Cherrier, Norman Poh, André Carlos Ponce de Leon Ferreira de Carvalho, Christophe Rosenberger, and Najoua Essoukri Ben Amara. 2019. Adaptive Biometric Systems: Review and Perspectives. *ACM Comput. Surv.* 1, 1, Article 1 (May 2019), 37 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The term *biometrics* generally refers to biological, morphological and behavioral characteristics of human beings. However, it is increasingly associated with automated techniques enabling the authentication of the identity of individuals based on their characteristics. Usually, the identity claim of individuals is verified based on what they own (*e.g.*, a token, a card) or what they know (*e.g.*, a password, a PIN code). However, for the biometric recognition, this verification is based on what the individual is/does, namely the person's biometric characteristics, such as fingerprint, hand signature or voice. Biometrics is often regarded as one of the most important solutions to security problems involving logical access control (*e.g.*, a computer, a network) and physical access control (*e.g.*, buildings, airports; electronic commerce; telephony; and identity management systems). According to [104], the global biometrics market will increase from $2 billion in 2015 to $14.9 billion by 2024.

The characteristics used for recognition should have some properties [37] to be considered useful: *universality* (everyone has the characteristic), *distinctiveness* (the characteristic allows to differ one user from another), *permanence* (the characteristic should not change over time for the same user), *collectability* (the characteristic can be measured quantitatively). Additionally, biometric authentication systems should take into account [37]: *performance*, *acceptability*, *circumvention*.

In the literature, the performance of biometric authentication algorithms is mainly evaluated based on two types of errors: *false match*, when an impostor is wrongly recognized as a genuine user, and *false non-match*, when a genuine user is wrongly rejected. False matches can come from the lack of *distinctiveness* of the biometric characteristic, while false non-matches can come from sample acquisition noise (*e.g.*, dirty fingerprint reader), unrepresentative biometric references (*e.g.*, a biometric reference of the frontal face may not match a query of the user in a different face position) or template aging. Template aging can occur when the biometric characteristic undergoes changes over time (*e.g.*, the typing rhythm on a keyboard can be improved with experience). All of these factors result in intra-class variability for the genuine class, which can negatively affect the biometric system recognition performance. A simple, although costly, way to deal with that is to resort to a human operator. For example, template aging can be handled by performing periodical enrollment sessions. Another solution is to use adaptive biometric systems.

Adaptive biometric systems automatically adapt the biometric reference over time. They can be used to either take into account template aging [77, 81, 96] or to improve unrepresentative biometric references [77, 116]. In several implementations, adaptive biometric systems exploit query samples (used for recognition) to adapt the biometric reference over time, reducing errors due to intra-class variability. This approach has many advantages since no operator is needed anymore. Adaptation is totally transparent to the users who are not explicitly asked for re-enrollment sessions. Adaptive biometric systems represent a relatively new research area in biometrics, with numerous open challenges. For instance, since the update procedure is automatic, the system may be subject to adversarial attacks which could introduce impostor patterns into the biometric reference. Therefore, some systems choose to only rely on highly confident samples.

This paper contributes to the field of adaptive biometric systems in the following ways:

- It provides the most up-to-date and comprehensive survey on adaptive biometric systems. The latest review we are aware of was published in 2012 [77] by Poh et al. and significant advances have been proposed ever since.
- This survey stresses the links and differences between existing works and emphasizes the lack of a common vocabulary between the published works in the literature. For this purpose, a formalization of adaptive biometric systems is proposed.
- The current paper also introduces a new taxonomy of adaptive biometric systems, which promotes a modular view in which each module or component can be independently analyzed. Moreover, a generic workflow of a biometric adaptation process is presented. Throughout the paper, a discussion of previous adaptation strategies is presented within the proposed taxonomy.
- An extensive discussion on several aspects of evaluation methodologies for adaptive biometric systems is presented. Diverse methodologies have been adopted in the literature and this paper attempts to provide a fair discussion of each aspect involved in the evaluation of adaptive biometric systems.
- This paper has a key importance for current and future research as it presents existent solutions, gaps and an up-to-date discussion of challenges involved in adaptive biometric systems.

The rest of the paper is organized as follows: Section 2 presents a formalization for adaptive biometric systems, along with the discussion of some biometric recognition errors; Section 3 decomposes adaptive biometric systems into modules which form the basis of a taxonomy to describe and compare current adaptive biometric systems; Section 4 discusses the evaluation methodology for adaptive biometric systems, including the presentation of datasets, metrics and other aspects involved in their evaluation; Section 5 points out several open research challenges for adaptive biometric systems. Finally, Section 6 presents the main conclusions.

2 ADAPTIVE BIOMETRIC SYSTEMS

As described in [37], a *biometric system* is a pattern recognition system that acquires a *biometric query sample* from a *claimant* and extracts its *biometric features* from the acquired sample. Next, the biometric system compares these biometric features to the *biometric reference* (also known as *model* or *template*) from the *claimed identity*, previously stored in a biometric database [37].

According to previous studies, biometric features may change over time [96]. Consequently, the biometric reference may no longer represent the current biometric features of the reference user. This phenomenon is known as *template aging* [39]. As a result, the recognition performance of the biometric system can degrade over time. An *adaptive biometric system* adapts the user reference to deal with template aging [77, 96]. This section presents the terminology adopted in this paper, discusses the need for adaptation, states the problem handled by adaptive biometric systems and some inherent aspects.

2.1 Biometric systems

A standard biometric system comprises two main phases: the *enrollment* and the *test/recognition*. During the enrollment, defined by Equation (1), the system receives a set of enrollment samples $\mathcal{E}_j$ for each user $j \in \mathcal{J}$ and outputs the biometric reference ref_j, where $\mathcal{J}$ is the set of user indexes registered in the biometric system. The enrollment is performed for all registered users and each biometric reference is stored in the biometric database $\mathcal{R} = \{ref_j \mid j \in \mathcal{J}\}$.

$$ref_j \leftarrow enroll\left(\mathcal{E}_j\right) \tag{1}$$

During the *test/recognition*, also known as operational phase, the system receives a biometric query sample $\mathbf{q}$ and returns the identity label of the recognized user. A query is a biometric sample acquired to perform recognition. The *test/recognition* can operate in two modes: *verification* or *identification* [39].

In the *verification* mode, defined in Equation (2), the query $\mathbf{q}$ is compared to the biometric reference ref_j of a claimed user with index j given a set of parameters θ_j^{verify}. The output is obtained from a *classification algorithm*, which returns the predicted label $label^p$ for the biometric query: *genuine* or *impostor*. The set θ_j^{verify} refers to the parameters adopted for the classification algorithm. Some implementations output a *score* from the comparison of a query $\mathbf{q}$ to the biometric reference ref_j and, afterwards, return the class label by comparing this *score* to a *decision threshold* value. In this case, the *decision threshold* would be an element in the set of parameters θ_j^{verify}. Other classification algorithms may need additional parameters, like the kernel parameters required by support vector machines [102].

$$label^p \leftarrow testVerify\left(ref_j, \mathbf{q} \mid \theta_j^{verify}\right) \tag{2}$$

In the *identification* mode, defined in Equation (3), the query $\mathbf{q}$ is presented to the biometric system, which outputs a set of user indexes $\mathcal{U}_{id}$ using the set of parameters $\theta_j^{identify}$, such that $\mathcal{U}_{id} \subseteq \mathcal{J}$. The set $\theta_j^{identify}$ refers to the parameters of the classification algorithm used, as in the case of the verification mode (*e.g.*, *decision threshold*). Note that $\mathcal{U}_{id}$ can be an empty set $\emptyset$ when the query is classified as an *impostor*, *i.e.*, the subject is unknown to the system.

$$\mathcal{U}_{id} \leftarrow testIdentify\left(\mathcal{R}, \mathbf{q} \mid \theta_j^{identify}\right) \tag{3}$$

2.2 On the need to perform adaptation in biometric systems

Although biometric systems represent a robust method to authenticate users, it has been reported that their recognition performance can degrade over time. As shown in previous studies, biometric features can change [96] and, consequently, the biometric reference may no longer represent the biometric features of the enrolled user. This phenomenon is known as *template aging* [39]. In machine learning, the term *concept drift* is often used to refer to changes in the profile of the data distribution [120]. In biometrics, drift is caused by numerous sources of variability. Furthermore, previous experimental results have shown that this variability can be user dependent [52, 54, 56, 67, 72, 88].

There are two main sources of variability over time:

- *Enrollment/changing conditions.* The model may not accurately represent the user characteristics when a limited number of samples is available during the enrollment stage. Moreover, these limited stored samples are usually not able to cover all possible conditions that will be encountered during the recognition phase. Aspects like illumination, humidity, noise, movement and portability of the device can vary between enrollment and recognition phases [78]. Another source of variability is the use of devices with distinct characteristics for enrollment and recognition (*cross-device matching*). This can occur, for example, in face recognition, when the quality of images produced by high-resolution cameras and webcams can be very different [71]. In fingerprint recognition, this can occur when matching two samples collected with thermal and optical fingerprint sensors. Self occlusions (*e.g.*, make up) and occlusions due to use of accessories (*e.g.*, body-piercing ornament, jewelries, glasses) can also introduce intra-class variability. Additionally, the interactions between a

user and each sensor can be different. In keystroke dynamics, typing on different keyboard layouts can produce distinct keystroke dynamics [39]. Emotion and health can also impact the recognition performance of a biometric system, especially in the behavioral modalities. Emotional states, such as happiness, anger and stress can impact the speech. Adaptive systems have also been used in the context of liveness detection; Rattani and Ross [93] improved the performance of their liveness detector by adding novel detectors for new kinds of spoofing material in order to automatically retrain their liveness detection system.

- *Time/aging*. Both physical and behavioral biometric modalities are subject to changes related to time/aging. Physical modalities are subject to injuries, wrinkles, speckles, weight loss and gain. Moreover, illnesses and their associated treatments can impact speech and fingerprint. Behavioral modalities are also subject to changes. For example, in keystroke dynamics, users may change their typing rhythm over time [59].

When genuine users are increasingly rejected by a biometric system, they can become annoyed, which negatively impacts the usability of the system. Periodical re-enrollment of the users can be a solution, although costly.

In short, intra-subject variability can increase the risk that the biometric system wrongly rejects a genuine attempt, thus increasing false non-match error. Two alternatives to decrease the impact of intra-class variability are using multi-modal biometric systems and adopting soft biometrics.

Multi-modal biometric systems use multiple biometrics [99] to reduce the overall system error that can occur at different levels (sensor, characteristics, score, rank or decision). Nevertheless, the configuration of the parameters of these systems can become complex, increasing their cost. The fusion may also be inconvenient to the user, since it can increase the overall authentication time.

Another alternative is to use soft biometrics [38]. Different from classical biometrics, soft biometrics [38] can improve biometric system performance by using characteristics that, even not being unique nor permanent to distinguish two individuals, can support the recognition decision. Examples are gender, age, ethnicity, skin or hair color. Soft biometrics have been successfully applied to different biometric modalities, like face recognition [18] and keystroke dynamics [35].

Despite the performance increase obtained by these alternatives, they are still subject to template aging. For example, in a multi-modal system using fingerprint and face recognition, when biometric features for both biometric modalities change, the recognition performance can degrade over time.

Deep learning, a trending topic nowadays, has been successfully applied to several biometric modalities [109]. One popular approach to using deep learning in biometrics is to employ a trained neural network as an encoder [103]. However, even though deep learning-based systems have obtained promising results, adaptation is still required. Indeed, after accumulating a certain amount of changes, any classifier would lose recognition performance. For instance, when the appearance of the biometric trait changes due to aging, the biometric reference encoded by the feature vector will no longer become representative of the identity. It is even more critical for behavioral modalities, which are subject to a higher degree of changes over time than physical modalities [30]. In short, adaptation is still needed to compensate for a myriad of unexpected or abrupt changes which may not be accounted for or represented in the enrollment set.

2.3 Fundamental issues behind adaptive biometric systems

In view of the changes in the biometric features previously discussed, there is a need for biometric systems able to adapt biometric references over time. These systems are known as adaptive biometric systems. In order to automatically adapt the biometric reference, an adaptive biometric system can use its database of biometric references ($\mathcal{R}$) and the unlabeled query samples collected over the use of the system. Some previous studies considered the availability of additional data to support

adaptation, such as session period [57] or true labels of the queries [112]. However, these additional data may not be available in a practical application, as discussed in Section 4.

Since an adaptive biometric system uses labeled data (*i.e.* enrollment samples) and unlabeled data (*i.e.* collected query samples), it can be seen as an instance of *semi-supervised learning* [14, 118]. In fact, implementations of adaptation strategies, like Self-update [98] and Co-update [97], are directly based on semi-supervised approaches: Self-training and Co-training [14, 118], respectively. These approaches can update the biometric reference by adding potentially novel patterns that can represent the genuine user (they are discussed in Section 3). Another application of adaptive biometric systems is to improve the biometric reference when there is a limited amount of training samples [116].

However, a key point has to be considered: adaptive biometric systems not only add patterns, but they can also discard outdated patterns from the biometric reference. It means that, due to changes over time, enrollment samples previously regarded as genuine samples may not represent the genuine user anymore. In this sense, adaptive biometric systems may differ from several semi-supervised learning applications, where the patterns corresponding to labeled data usually remain unchanged over time. In biometric systems, labeled data correspond to the enrollment samples. Due to variations in the biometric features over time, these labeled samples may no longer accurately represent the genuine user.

Within the machine learning community, this phenomenon can be related to concept drift observed in streaming data [120]. The term *concept drift* is often used to indicate changes in the profile of the data distribution. Indeed, some studies have considered the sequence of queries as a pool [31] or as a biometric data stream [62, 70].

2.3.1 Some insights on why adaptive biometric systems work. Some assumptions from semi-supervised learning are expected to remain valid for adaptive biometric systems, such as the smoothness and the cluster assumptions [14]. If two samples are close and can be placed on the same cluster, they are likely to have the same label.

In semi-supervised learning, the use of unlabeled data has been a subject of criticism. Indeed, the performance can degrade if these data are used in generative models [14, 15]. It is important to highlight, however, that other approaches may need additional studies to confirm that they suffer from the same performance degradation observed in generative models, as discussed in [15].

Although those studies were mainly targeted on generative models in semi-supervised learning, a related question could be raised for adaptive biometric systems: why should adaptation work on biometric systems? Usually, only queries with a certain level of similarity are used to adapt biometric references; this implies that gradual changes can be captured, something that can be expected from most changes related to time or aging. Researches on adaptive biometric systems have mainly focused on this kind of adaptation.

Abrupt changes, on the other hand, are hard to be identified and, usually, will not be captured by most adaptive biometric systems. An abrupt change could happen if, for instance, the user accidentally injures one of their fingers: a fingerprint recognition system might fail to recognize this user later. Another possible source of abrupt changes could be different acquisition conditions, such as those listed in Section 2.2. This topic requires further work and is listed as an open challenge in Section 5.

Let's illustrate these gradual changes and how a biometric system can handle it with some keystroke dynamics data. In keystroke dynamics, individuals are recognized by their typing rhythm. This biometric modality is subject to a high rate of changes over time and it disposes of several publicly available data captured over time (see Section 4.1). Related conclusions can also be obtained from other modalities, such as the changes in face recognition [90]. Figure 1 presents a summary of

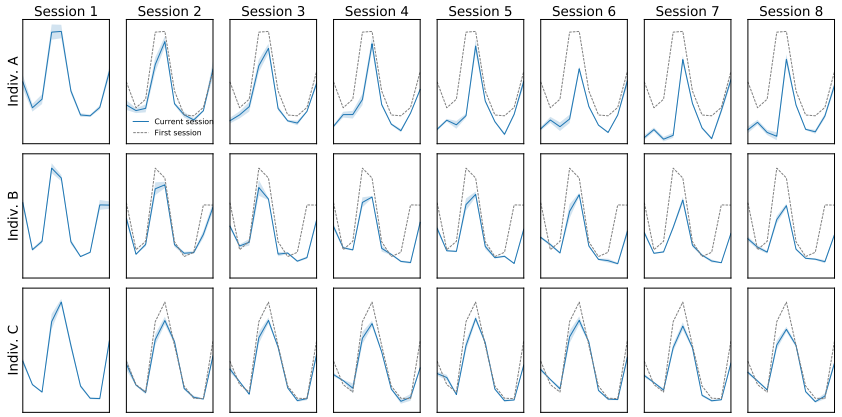
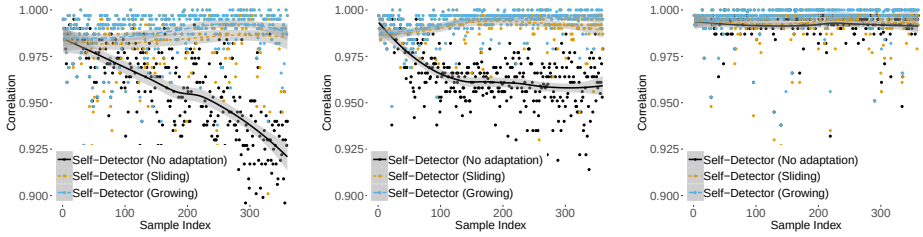


Fig. 1. Summary of raw keystroke data of three individuals from the CMU dataset over time. Each session-plot shows the average of its keystroke samples and compares it to the data in the first session.



(a) Indiv. A. (b) Indiv. B. (c) Indiv. C.

Fig. 2. Correlation over time between keystroke dynamics samples and owner's biometric reference. Individuals A and B can benefit significantly from adapation in comparison to individual C.

the raw biometric data of three individuals from the CMU keystroke dataset [43]. Each row of these plots shows a sequence of line charts for a different user and each of them displays a summary of the keystroke data acquired at a specific acquisition session; thus, each sequence illustrates how keystroke data changes over the sessions. By looking at these three individuals, we observe that biometric data can change in different ways; A and B were subject to a higher degree of changes over time than C.

Figure 2, based on the idea of [62, 67], illustrates the impact of adaptation over time for three individuals. The maximum correlation between each genuine sample and the biometric reference is shown for different biometric systems. The black data shows correlations for a biometric system without adaptation (no adaptation) whereas the blue and the yellow data shows the correlations for two different adaptive biometric systems (Growing and Sliding). The correlation can decrease for a non-adaptive biometric system, while the adaptive biometric systems managed to keep it at a higher value over time. Individual C shows almost no changes over time, suggesting that adaptation may not be required for that user. The use of adaptation has not strongly impacted the recognition performance for this individual. A higher correlation over time means that the genuine biometric data has not drifted away from the user reference. Adaptive strategies can keep this correlation at a higher rate over time, mitigating the impact of changes in the biometric features.

In practical applications, gradual variations of the biometric features are expected to be more common than abrupt changes. Based on this assumption, adaptive biometric systems could handle most of the cases and avoid or delay the need for re-enrollment of the users.

2.3.2 Situations that can arise from adaptive biometric systems. Adapting biometric references from unlabeled queries is a challenging task: an adaptive biometric system has to maintain the biometric reference updated, while avoiding the inclusion of impostor patterns in the genuine reference. This section discusses some key situations and contributes to highlight the main challenges of adaptive biometric systems.

In the enrollment process, which is standard for any biometric system, the biometric reference is computed, as shown in Equation (1). The ability of the biometric reference to properly model the genuine user depends on (i) the quality and (ii) the representativeness of the enrollment samples as well as on (iii) the overall performance of the recognition algorithm. Different cases are expected:

- *The biometric reference no longer represents the subject's biometric trait.* This may be due to a limited amount of enrollment samples. Adaptation can address this issue by updating the biometric reference. This is a case where adaptation is not used to deal with changes over time. It is important to note, however, that it may be hard to automatically update the reference if the owner is systematically rejected.
- *The biometric reference properly represents the subject's biometric trait and there is no variability.* There is no need for adaptation in this case.
- *The biometric reference properly represents the subject's biometric trait and there is variability over time.* Some papers have shown that the biometric features may change in different ways, depending on the user [67, 72]. The use of an adaptation mechanism is likely to improve the biometric reference, avoiding performance degradation over time.

After the enrollment, another standard process is the test/recognition, which is the verification of the identity of the claimant by comparing the provided query to a biometric reference, as shown in Equation (2). If the provided query appears significantly different from the biometric reference, the claimant is rejected. This process has a key role in an adaptive biometric system, since queries that can be used for adaptation are received during the test/recognition phase. Four cases are expected:

- *Genuine claimant was rejected.* The biometric reference does not properly model its owner and updating the biometric reference could avoid further false non-matches. However, since the user was rejected, it is very unlikely that the query will be selected for adaptation. This particular case may occur due to abrupt changes.
- *Genuine claimant was accepted.* Although the biometric reference properly models its owner, updating the biometric reference with the new query may add new patterns resulting from a gradual change. There is a high probability that this query is selected for adaptation.
- *Impostor claimant was rejected.* The biometric reference correctly detected an impostor. If the adaptive biometric system is also able to model the impostors, it can use the current query for that [63].
- *Impostor claimant was accepted.* This is one of the hardest cases to handle. It is likely that the query will be used for adaptation, reinforcing the error. It is similar to a problem that affects Self-training [118]. In fact, this represents an important challenge in adaptive biometric systems, as described in Section 5.

Different approaches have been employed to mitigate the previous problems, such as the use of multi-gallery adaptation mechanisms (Section 3.5.3). Another approach is to maintain two separate models, namely a genuine and an impostor model [63].

Table 1. Recurrent terminology adopted throughout the paper.

Terminology	Meaning
$\mathcal{J}$	Set of user indexes registered in the biometric system
$j \in \mathcal{J}$	A registered user index
ref_j	Biometric reference of a user j
$\mathcal{R} = \{ref_j \mid j \in \mathcal{J}\}$	The set of stored biometric references (biometric database)
q	A biometric query sample
$label^p$	Predicted label for the biometric query: *genuine* or *impostor*
θ_{verify}	Set of parameters for the test/recognition process (verification mode)
$\theta_{identify}$	Set of parameters for the test/recognition process (identification mode)
$\mathcal{A}$	Set of samples for the adaptation process
θ_{adapt}	Set of parameters for the adaptation process

Furthermore, the implementation of the adaptation process also plays a key role. Apart from the issues mentioned in this section, the adaptation mechanism also impacts the quality of the updated biometric reference. As discussed throughout Section 3, some adaptation mechanisms only add new patterns, while others can also discard outdated patterns. Since the users may naturally change their biometric features in different ways, there is no common best choice. Indeed, as suggested in [62], the optimal adaptation strategy may differ from one user to another.

All in all, the updated reference is expected to better represent the genuine user. As a result, the user is less often rejected and impostors are less often accepted. Conversely, in the worst case, the updated reference drifts from its owner and better represents the rest of the world: the genuine user is more often rejected and the impostors are more often accepted. Since the adaptation is usually unsupervised, it is hard to guarantee that the system perform no worse than before. Monitoring the error rates over time could be an option [106], although it may not be feasible in several scenarios. An alternative is to use active learning [1], where the biometric system would select some queries to be manually labeled, which is applicable to some scenarios.

2.4 Adaptive biometric systems

After the discussion of fundamental issues behind adaptive biometric systems, this section presents a proposal to formally define an adaptive biometric system. In short, it can be understood as a standard biometric system with an additional phase: the *adaptation process*.

In the adaptation phase, the *adapt* process, as specified in Equation (4), adapts the biometric reference $ref_{j_{(t)}}$ using a set of biometric samples for adaptation $\mathcal{A}$ given a set of adaptation parameters θ_j^{adapt}. The output of the adaptation process is the *adapted biometric reference* $ref_{j_{(t+1)}}$.

$$ref_{j_{(t+1)}} \leftarrow adapt\left(ref_{j_{(t)}}, \mathcal{A} \mid \theta_j^{adapt}\right) \tag{4}$$

The set of samples used for adaptation, $\mathcal{A}$, is collected during the system operation. Usually, it only contains samples classified as genuine by the test/recognition process. As discussed later, some studies only include samples classified as genuine with high confidence in this set [24, 98]. For such, an additional *adaptation threshold*, that is more stringent than the *decision threshold*, can be used. In this case, the *adaptation threshold* would be an element in the set of parameters for adaptation θ_j^{adapt}. Depending on the adaptation strategy, there may be different parameters, as discussed in the next sections of this paper.

Many *adaptation strategies* consider that the biometric reference ref_j is composed of several biometric samples/templates (Section 3.1). This set of samples/templates is sometimes referred to as a *gallery* [6, 29, 57, 86]. In line with this concept, adaptation can be defined as the addition and removal of samples/templates from a gallery.

The adaptation process can be performed either *online* or *offline* [77] (Section 3.4). In the online adaptation, the process is executed after each query sample is recognized by the biometric system. Basically, the adaptation process is triggered every time the test/recognition is performed. In the offline adaptation, however, instead of triggering the adaptation process after each query recognition, the system waits to store a batch of biometric samples in the set $\mathcal{A}$ before adapting the biometric reference.

In this paper, the behavior of adaptation is determined by the *adaptation strategy*, which relies on an *adaptation criterion* (Section 3.2) and on an *adaptation mechanism* (Section 3.5). In this line, an adaptive biometric system is composed of a *classification algorithm* and an *adaptation strategy*. Table 1 presents recurrent terminology adopted throughout the paper.

This survey focuses on adaptive biometric systems, which are able to automatically adapt the biometric reference over time. They are sometimes referred to as *template update* in the literature. Next sections focus on adaptation strategies, evaluation methodology and future work opportunities for adaptive biometric systems.

3 STRATEGIES TO ADAPT THE BIOMETRIC REFERENCE

A number of adaptation strategies have been proposed in the literature. To the best of our knowledge, this survey presents the most comprehensive and up-to-date collection of adaptation strategies. No previous review [16, 77, 78, 105] in the field of adaptive biometric systems have provided such an extensive and complete overview of the field. This section presents adaptation strategies found in the literature based on five distinctive aspects, as shown in the taxonomy in Figure 3:

- *Reference modeling*. How the biometric reference is modeled.
- *Adaptation criterion*. The criterion chosen to trigger the adaptation mechanism.
- *Adaptation mode*. The method employed to assign labels: supervised or semi-supervised.
- *Adaptation periodicity*. The periodicity in which the adaptation process is applied: online or offline.
- *Adaptation mechanism*. How adaptation is performed (when the adaptation criterion is satisfied).

These five aspects are further divided into additional categories, composing the complete taxonomy. Previous reviews have presented some taxonomies [77, 84]. However, they are not as extensive and up-to-date as the proposal in this paper. Moreover, those taxonomies have not adopted a modular view of the adaptation strategies as proposed in this paper. To this end, a generic work-flow of a biometric adaptation process is shown in Figure 4. Next subsections present previous work within the proposed taxonomy.

3.1 Reference Modeling

Biometric reference modeling strategies can impact how to adapt/update it over time. For instance, a speech signal can be represented as Mel-Frequency Cepstral Coefficient (MFCC) features, whose density is modeled using a Gaussian Mixture Model. Thus, the resulting reference is a statistical model [94] storing the biometric features. Another example is when the k-nearest neighbor (k-NN) [8] algorithm is used by a biometric system. In this case, its reference is a set of samples. A related concept is adopted when the biometric reference consists in a set of detectors [67]. Each type of reference modeling may need distinct adaptation mechanisms. Overall, three main categories of biometric references can be found in previous studies on adaptive biometric systems:

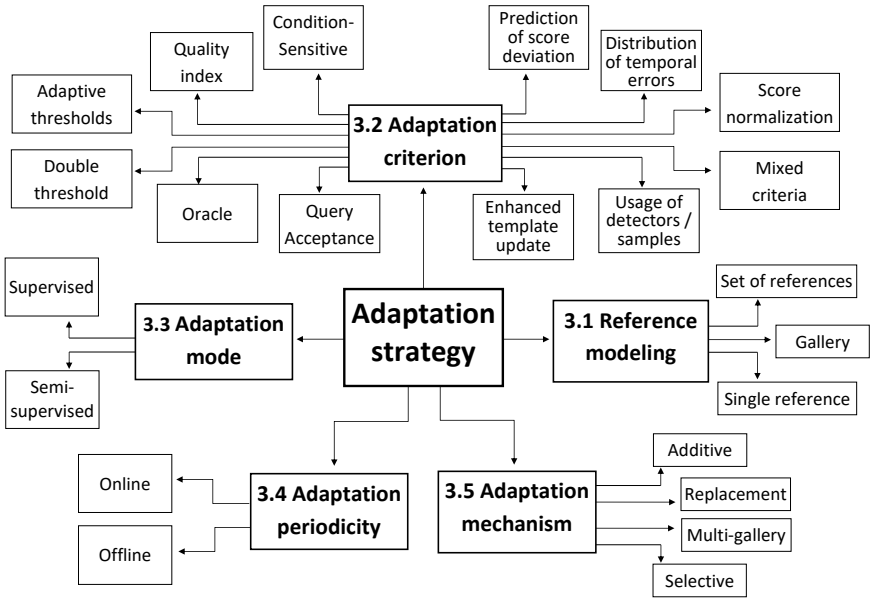


Fig. 3. Proposed taxonomy of adaptive strategies based on five distinctive aspects: reference modeling, adaptation criterion, adaptation mode, adaptation periodicity and adaptation mechanism. Each aspect is numbered and refers to the respective section in the paper.

• *References containing a single sample/template.* The biometric reference [32] can be a single good quality capture acquired at the enrollment phase. Although this category has been used for physical modalities, it may not be reliable for behavioral biometrics where a single sample is unlikely to capture enough variability usually present in behavioral modalities. Nevertheless, some recent papers on keystroke dynamics have shown adaptation using a single sample during the enrollment phase [54–56].

• *References built from several samples/templates.* Several samples are acquired during the enrollment phase and stored in a *gallery*. In some studies, each sample is known as a *detector* [67]. Using galleries in adaptive biometric systems is a very common approach as shown in the next sections.

• *Set of references.* Several references per user are organized to represent different aspects of the biometric data [46]. Other examples are the biometric references used in [29] and [63], which contain two sub-references to support recognition and adaptation.

3.2 Adaptation Criterion

The adaptation criterion determines if adaptation should be performed or not. Several criteria have been proposed in the literature:

• *Call for an oracle.* The decision to use a query for adaptation is taken by an oracle. It can be, for instance, a human operator [108, 112].

• *Query acceptance.* Each accepted query is used to adapt the reference [41, 113].

• *Double threshold.* An *adaptation threshold* is adopted in addition to the *decision threshold* already used for the recognition process. Query samples that meet the *adaptation threshold* are used for

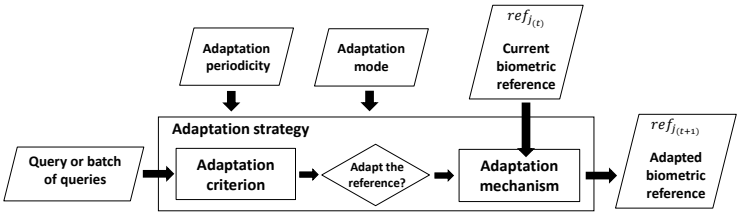


Fig. 4. Generic work-flow diagram of a biometric adaptation process, where the biometric reference ref_j of user j can be updated by the adaptation strategy.

adaptation. As the *adaptation threshold* is usually more stringent than the decision threshold [80], only highly confident queries are used for adaptation.

• *Adaptive thresholds.* The adaptive thresholds method [57] extends the double threshold one by updating both decision and adaptation thresholds over time, allowing them to follow the intra-class variation of user characteristics over time.

• *Quality index.* The quality index of the query can be a criterion to decide whether a given query should be used for adaptation [60, 73].

• *Condition-sensitive.* Adaptation is performed if conditions not present in the biometric reference are observed during operation, such as a different pose or illumination [61], or, for example, to detect new materials in a liveness detector [93].

• *Prediction of score deviation.* This criterion analyses the scores of the biometric system to estimate when the biometric reference should be adapted [12].

• *Distribution of temporal errors.* In an operational scenario, false non-matches can bring useful information to the system [106]. For example, a continuous sequence of false non-matches could mean that the biometric reference has aged and, consequently, it should be updated.

• *Mixed criteria.* It is used under a multi-modal biometric system. For example, if the system uses a modality with high intra-class variability and another modality less affected by intra-class variations [97], adaptation can be performed if the number of non-matches by the first modality reaches a given threshold, while the second modality accepts the user. It can also occur at the feature level [42] by using invariant features to confirm the adaptation of the variant features.

• *Enhanced template update (ETU).* A system can be designed to model an individual with two sub-references: a genuine reference, which models the biometric features of the target individual, and an impostor reference, which models the features of everyone else. The genuine reference is adapted using queries accepted as genuine, while the impostor reference is adapted using the rejected query samples. These sub-references can be used in different ways to support verification and adaptation [63].

• *Usage of detectors/samples.* It relies on the concept of checking the usage of detectors (*i.e.* biometric samples) from the biometric reference for matching to discard unused detectors over time. Some variations were proposed [65–67]. In Usage Control/Usage Control R/Usage Control 2 (see Section 3.5), adaptation occurs if some detectors have not been recently used. Usage Control S additionally checks whether at least two detectors match the input query.

• *Score normalization.* As discussed in Section 2.1, some systems output a score from the comparison between a query sample and a biometric reference. Based on this score, a threshold is applied to both output the label (genuine or impostor) and to decide whether adaptation should occur. Score normalization [76] refines the output score and, consequently, allows a better choice of thresholds. A preliminary study on the use of score normalization for supervised adaptation to handle different

acquisition conditions is shown [73]. Later, the use of score normalization in adaptive biometric systems was further studied in [70], considering a biometric data stream context.

The first criterion requires an oracle to tell when adaptation should be performed [108] and it is not always feasible. Query acceptance is a simple criterion that avoids this problem [41]. However, it is prone to allow the inclusion of wrongly classified query samples into the genuine biometric reference. An alternative to deal with this problem is the double threshold [80], which uses an additional threshold for adaptation. Nevertheless, the double threshold criterion usually just captures little variability, since only query samples with a high probability of belonging to the genuine user trigger the adaptation process. Although these methods can decrease the risk of wrongly including impostor samples in the genuine biometric reference, the expected performance gain due to the adaptation strategy is likely to be limited. Adaptive thresholds [57] can be considered an improvement over the double threshold as it adapts the thresholds over time [54–57]. Quality-index may also be used to only add high-quality data to the biometric reference [60, 73].

The condition-sensitive criterion provides a way to avoid including redundant information into the biometric reference. This is because it only adds new samples if novel conditions are identified during the operation of the biometric system [61].

It can also be possible to predict when adaptation should be performed by checking the score deviation [12]. However, the prediction may not be accurate if the biometric features from the users start to change in a different way over time. The distribution of errors over time may also indicate the need to adapt the biometric reference [106]. Nevertheless, it assumes that false non-matches can be reliably measured during system operation. For example, in border control, customers who have a refused entry would have to go to a separate queue for manual identity verification. Therefore, closely monitoring the error over time constitutes a viable criterion for adaptation.

Using multiple sources to support the adaptation criterion is observed in the mixed criteria [97] and the enhanced template update [63]. The former work uses multiple biometric modalities in a multi-modal system, while the latter stores a genuine and an impostor model to support the decision to whether or not perform adaptation.

The usage of detectors for matching can also provide information to decide whether adaptation should be started. Various ways of using this information have been proposed [65–67].

Score normalization is an alternative to refine the output score in adaptive biometric systems [73, 76]. As a result, a better threshold choice can be done, improving the performance of the adaptation criterion. Previous work has applied score normalization to several adaptation strategies in a biometric data stream context [70]. Applying score normalization requires additional data, either a development or a cohort database depending on the normalization procedure.

As discussed in this section, there are several criteria that can be adopted to decide whether adaptation should be performed or not. They rely on different aspects, such as score, quality, errors and usage. However, they are still prone to adversarial attacks, which could introduce impostor patterns into the genuine biometric reference [6, 7]. This a topic not deeply explored in the literature which is further discussed in Section 5. To summarize the discussion so far on adaptation criteria, Table 2 highlights their advantages and drawbacks.

3.3 Adaptation mode

Query samples are usually unlabeled. However, in some cases, true labels could be received some time after the biometric system has classified them, similarly to data stream mining applications [120]. When query samples are unlabeled, semi-supervised adaptation is performed. Conversely, when they are labeled, supervised adaptation techniques can be used. This section

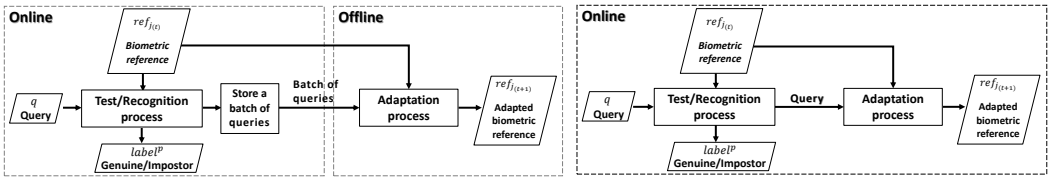
Table 2. Comparison of adaptation criteria.

Criterion	Advantages	Drawbacks
Call for an oracle [108]	- The method is secure; - Uses only close genuine biometric samples from the biometric reference.	- It is manual.
Query acceptance [113]	- The method is simple and allows automatic adaptation.	- Can include characteristics of wrongly accepted impostors in the genuine biometric reference.
Double threshold [80]	- Can reduce the inclusion of impostors samples in the biometric reference by an additional (more stringent) adaptation threshold.	- It is only able to capture little variability.
Adaptive thresholds [57]	- User-specific adaptation of the thresholds over time.	- The initial thresholds must be well chosen to obtain good performances.
Quality Index [60, 73]	- Avoids the use of low quality samples in the adaptation; - Can replace low quality data acquired in the enrollment procedure.	- Need to define the quality index, which can be modality dependent.
Condition-sensitive [61, 93]	- Excludes redundant information and can potentially reduce the size of the reference, saving computer resources.	- Sensitive to the initial samples in the reference as well as the updating threshold.
Prediction of score deviation [12]	- Prediction of the moment to update the biometric reference.	- If the pattern in which the biometric features change over time, the prediction may not be accurate.
Distribution of temporal errors [106]	- Monitors the actual error to mitigate it.	- Requires a way to measure false non-matches over time.
Mixed criteria [97]	- Uses additional information from multiple biometric modalities.	- Requires more than one biometric modality, increasing costs.
Enhanced template update [63]	- Combines a genuine and an impostor gallery to support both test and adaptation.	- Classification errors may poison both galleries.
Usage of detectors/samples [65–67]	- Keeps the biometric reference updated by the patterns most frequently and recently present in the queries.	- May remove true user patterns from the biometric reference if they are not frequently present in the queries.
Score normalization [70, 73]	- Refines the output score for a better threshold choice.	- Requires additional data to normalize scores (a development or a cohort database depending on the normalization procedure).

briefly describes them as two adaptation modes for adaptive biometric systems: supervised and semi-supervised adaptation.

• *Supervised adaptation.* It is usually easier than semi-supervised adaptation since it uses true labels of the query samples for adaptation. These labels are provided by an oracle, also known as an operator in this context. It has been extensively studied in the literature [20, 21, 25, 111]. The samples can be obtained in different ways. For example, several enrollment sessions can be applied to each user [111]. The newly acquired samples at each enrollment session are labeled and can be used to adapt the biometric reference. Of course, this approach can be time-consuming and expensive as it requires individuals to participate in several enrollment sessions. Moreover, an operator must supervise these enrollments to avoid errors. Another method to obtain labels for supervised adaptation is by manually labeling the captured data when the authentication system is in use, *e.g.*, in an operator-assisted face recognition system [108]. However, this approach is not applicable to many contexts of application, especially when the operator or supervisor is not available.

• *Semi-supervised adaptation.* It is a more realistic scenario [40, 57, 63, 75, 100], where the labels are provided by the biometric system automatically. Note that, in this case, however, the obtained labels can be wrong. It is semi-supervised as the system has the labeled samples from the initial



(a) Delayed/Offline adaptation. (b) Real-time/Online adaptation.

Fig. 5. Delayed/Offline vs Real-time/Online adaptation. The latter one can be viewed as a simplified version of the former since the adaptation criterion takes the decision immediately after the recognition process.

enrollment and the unlabeled query samples. The concept is to automatically label the query samples in order to use them during the adaptation process. There are two main ways to perform it: *self-training* [89] or *co-training* [10]. Self-training is related to mono-modal authentication systems and uses samples classified by the same classifier to retrain it, while co-training is related to multi-modal authentication systems and consists of using the knowledge of one modality to support labeling the other one.

To summarize, semi-supervised methods automatically provide labels to the collected queries thanks to the classifiers, whereas supervised ones rely on an oracle. The main drawback with semi-supervised is that the predicted labels can be wrong in case of recognition errors. Semi-supervised adaptation is most often studied as it is a more realistic, although complex, scenario.

3.4 Adaptation periodicity

The adaptation process does not have to be performed each time a query satisfies an adaptation criterion. There are two main settings, as illustrated in Figure 5:

• *Offline/delayed adaptation.* Queries are collected and stored in a buffer before processing them in a batch procedure. It is a common approach in the literature, however, the choice of the adaptation frequency remains an open issue. Which is the best strategy to adopt: waiting until enough samples have been collected or waiting for the expiration of a specific delay? As discussed later in Section 4, the periodicity has been determined by the dataset division of sessions in previous studies. However, it is still an open question in practical application scenarios.

• *Online/real-time adaptation.* This setting systematically performs adaptation after the decision criterion is met (often taken after the acceptance of the query [25, 66]). As the process is iteratively done, query per query, it mainly fits the semi-supervised adaptation mode, where the adaptation system uses the label computed by the verification method on the selected query.

Offline adaptation has the advantage of a minimum performance impact during the recognition process, since no adaptation is done while the system is operating. The adaptation process can then be triggered when the system is not in use. Online adaptation, however, adds more processing time during the recognition process, as both recognition and adaptation processes are performed jointly. Nevertheless, it must also be noted that the online adaptation does not need to store a buffer, hence, it consumes less memory than the offline setting.

The adaptation periodicity can also affect the choice of the adaptation mechanism. Some mechanisms were designed for offline adaptation, such as Graph min-cut (see Section 3.5) that needs a buffer of queries to build a graph, as part of its adaptation process.

3.5 Adaptation mechanism

As previously discussed, an adaptation strategy is composed of various modules. The adaptation mechanism is the one that finally adapts the biometric reference. All adaptation mechanisms presented in this section are suitable for references that are composed by a set of templates/samples, sometimes named as a gallery [6, 29, 57, 86]. Overall, the adaptation mechanism basically adds and/or removes samples/templates from a gallery. The biometric reference is re-computed upon gallery modification. Four main categories of adaptation mechanisms exist:

- *Additive mechanisms* receive a set of samples and add all (or some) of them to the gallery;
- *Replacement mechanisms* receive a set of samples and add all (or some) of them to the gallery, but they also remove some samples from it;
- *Multi-gallery mechanisms* manage two (or more) galleries and they can also apply distinct adaptation mechanisms to each gallery;
- *Selection mechanisms* select the most important samples in a gallery to keep, avoiding the gallery to indefinitely increase its size over time.

The above adaptation mechanisms are presented in the next sub-sections, along with a discussion of their advantages and drawbacks.

3.5.1 Additive mechanisms. An additive mechanism is based on the concept of progressively adding new patterns to the biometric reference. This mechanism can encode a higher variability of the user data, which can consequently avoid false non-match due to genuine intra-class variability. One of the first attempts in this direction was proposed by Uludag *et al.* [111]. The proposed mechanism, called *augment-update*, adds a set of new samples to the user gallery. Their experiments assumed that this new set of samples was genuine. Subsequent works on the additive mechanism used the predicted labels instead of the true labels for adaptation. Some of them are described next.

- *Self-Update* (Procedure 1 of the supplementary material). As described in [90], it is an implementation of self-training [89] for adaptive biometric systems [98]. It has been extensively studied in the literature [2, 24, 27, 80, 96, 98]. The general concept is to add query samples classified as genuine to the gallery. Usually, only those samples that meet a genuine similarity score above a given *adaptation threshold* are added to the gallery. Hence, Self-Update is commonly implemented together with *double threshold*, as described in Section 3.2.

Another related adaptation mechanism is the *Growing window* [41]. Growing window works similarly to Self-Update, however, it does not use the additional *adaptation threshold*. It can also be understood that it assumes that both decision and adaptation thresholds are the same. As a result, all queries classified as genuine are added to the gallery.

Concerning the adaptation periodicity, in the literature, Self-Update is frequently applied in a scenario adopting offline adaptation, where a batch of queries is received for adaptation from time to time. Conversely, Growing window is usually applied in online scenarios, where the adaptation process is executed after each query is processed.

- *Graph min-cut for template update* (Procedure 4 of the supplementary material). It is an adaptation mechanism proposed by [86, 90], which uses the *max-flow/graph min-cut* algorithm [9]. This adaptation mechanism receives a batch of query samples and joins them to the current gallery of the user. Based on this data, a graph is generated, where each node represents a sample and each weighted link is a similarity score between samples. The graph-min cut divides the graph into two parts: source (genuine samples) and sink (impostor samples). The source represents the new gallery. The way that graph is generated implies that no sample in the initial gallery is removed during adaptation (all samples from the gallery are assigned infinite weight to the source/genuine node), justifying the categorization as an additive mechanism.

Table 3. Comparison of additive mechanisms.

Mechanism	Advantages	Drawbacks
Self-Update [98]	- Simple to implement.	- The adaptation threshold can be difficult to define: low values may imply in several impostor samples included in the gallery, while high values can prevent proper adaptation to genuine data.
Graph min-cut for template update [86, 90]	- Able to capture higher intra-class variability than Self-Update.	- The computations can become intensive.
Adaptation using Harmonic function [85]	- Obtain good performance even with few labeled samples.	- The computations can become intensive.

• *Adaptation using harmonic function.* A work from [85] proposes an adaptation mechanism using harmonic functions, which makes use of probabilistic semi-supervised learning introduced in [119]. Similar to the previous adaptation mechanism based on graph min-cut, this adaptation mechanism also receives a batch of query samples and joins them to the current gallery of the user. The joined set of samples is used to compute an adjacency matrix, which is then applied to obtain a harmonic function for the set of query samples. The obtained harmonic function is employed to determine which queries are added to the gallery.

Self-update refers to a category of adaptation mechanisms that uses only one classifier [90, 98]. It is vulnerable to the mistaken introduction of impostor samples in the gallery. Although this problem is faced by most adaptation mechanisms, its impact is worst in the case of additive ones, since the gallery keeps growing and no sample is removed. Of course, this could be avoided using a very high *adaptation threshold*. Nevertheless, this also means that only those genuine queries very close to the current reference would be accepted for adaptation. Since they are already close to the reference, they could not bring enough new information and larger changes would not be captured. This illustrates that the configuration of the adaptation threshold deeply impacts the performance of the adaptation mechanism. Considering the graph-based mechanism, its authors claim they can capture larger intra-class variabilities than Self-Update [90]. However, this mechanism as well as the one based on harmonic function [85] needs more computer resources than Self-Update, and the computations can become intensive, particularly if the gallery and the set of queries are large.

Additive mechanisms have the drawback of indefinitely increasing the size of the gallery, which could lead to problems in terms of memory usage. A possible way to mitigate it would be to use selection mechanisms described in Section 3.5.4 after the additive mechanism is executed. Table 3 summarizes the advantages and drawbacks discussed here.

3.5.2 Replacement mechanisms. Mechanisms of this family also add new samples to the gallery over time. However, they additionally can remove the samples to avoid the problem of indefinitely increasing the gallery size. Again, one of the first attempts following this concept is from Uludag *et al.* [111], which presented the *batch-update*. This mechanism receives a set of samples and uses it as the new gallery. Thus, the entire previous gallery is discarded. In their experiments, it was considered that the true label is provided for the new set of samples, which may not be a feasible assumption in practice. Several replacement mechanisms are presented in this section. All of them assume that the query or the set of query samples received as input are classified as genuine.

• *Sliding/Moving window* (Procedure 6 of the supplementary material). This mechanism was described in [41], though it can also be found under the name of First In First Out (FIFO) [20, 101]. It receives a set of query samples (the set can contain just one sample) and adds them to the gallery

by removing the same number of oldest samples, thus keeping the gallery size constant over time. Double threshold criterion can be used with this mechanism. As a consequence, only samples that obtain similarity score above a given *adaptation threshold* will be added to the gallery. Another related adaptation mechanism is adopted in [32, 53, 55], which works similarly to growing window until the gallery reaches a maximum size, when it uses sliding window for the adaptation. This mechanism can be adjusted to the users categories according to Doddington's Zoo classification by tuning specific parameters (reference size and thresholds) for each class of users [52].

• *Replacement based on MDIST and DEND.* Freni *et al.* [20] proposed replacement mechanisms based on the operating principle of MDIST and DEND clustering algorithms [111]. The general concept is to add a new query sample and remove another one from the gallery, thus keeping the same gallery size after the adaptation. For such, all possible gallery variations are evaluated (each time a different sample is removed). The scores among all samples are computed for each gallery variation. This process is also performed for the unmodified gallery. Then, the average score for each gallery is obtained. Based on this average score, the gallery is chosen according to one of the two strategies here: for MDIST, the gallery which has the maximum average score is chosen, whereas, for DEND, the chosen reference is the one corresponding to the minimum average score.

• *Least frequently used (LFU).* LFU was presented in [20, 101] and consists in adding the received query sample to the gallery and removing the least frequently used ones. It requires to maintain the number of times each sample is used to authenticate the user.

• *Least recently used (LRU).* LRU proposes to replace the least frequently used sample of the gallery by the new query sample [101]. A first method is to use a timestamp for each sample of the gallery, but it can be too expensive. The authors then suggest using the *clock algorithm*, a special case of the second-chance approach [101].

• *Extended replacement.* It computes a relevance attribute for each sample of the gallery based on its usage for matching and performs replacement based on it [101]. The sample with the lowest value for this relevance attribute is removed and the new query sample is added to the gallery.

• *Usage Control.* It is based on the concept of checking the usage of detectors (biometric samples) of the biometric reference for matching to perform adaptation. The more recently (and frequently) used detectors are kept in the biometric reference, while the remaining detectors are removed. Four versions are proposed: Usage Control, Usage Control R, Usage Control S and Usage Control 2 [65–67]. If the adaptation criterion for them is met, a detector (or a set of detectors) is removed from the biometric reference. In Usage Control/Usage Control R, the mechanism first selects those detectors less recently used. Among them, the least frequently used is removed. Usage Control S works similarly, but it has a more stringent adaptation criterion (at least two detectors should match the input query). Usage Control 2 can remove more than one detector since it removes all detectors not recently used. As an example, the algorithm for Usage Control/Usage Control R is given in Procedure 5 of the supplementary material.

• *Transfer learning-based.* In [13, 121], the authors presented adaptive mechanisms based on transfer learning [110] to update SVM classifiers. Given an SVM trained on the enrollment data, the adaptive mechanism is capable of adapting it using later acquired labeled samples. Note that these mechanisms do not use a gallery as the other ones presented here. However, as it replaces an older user model with a newer one (adapted using transfer learning), it is classified as a replacement mechanism in this paper.

As stated at the beginning of this section, a key advantage of a replacement mechanism is that it can avoid increasing the gallery indefinitely over time. The crux of maintaining the gallery size is that when a new sample is added to the gallery, another one has to be removed. This

section presented several ways to choose which samples are replaced. The simplest one is sliding window/FIFO [41, 101], which simply replaces the oldest sample(s). This mechanism assumes that the most recent samples are more representative, though it may not always be the case.

MDIST and DEND [20] can be computationally intensive if the gallery is large, since it requires to compute the scores among all samples for several gallery variations. Since MDIST keeps the gallery with the highest average score, the obtained gallery has less variability among the samples than the gallery obtained by DEND (which keeps the gallery with the lowest average score). The authors mention that MDIST is based on the idea of keeping samples that are similar to exploit common representative characteristics, while DEND is able to represent larger intra-class variability.

A technical report [101] presented three adaptation mechanisms that replace samples considering their usage, although none of them were experimentally evaluated. LFU replaces the most frequently used sample. If a sample is too frequently used for some time, it can be hard to be replaced later if it becomes unrepresentative of the current user data. Moreover, older samples tend to be more used, making the mechanism subject to replace newer samples over time, which may not be the most suitable choice. LRU then replaces the least frequently used, but it may be expensive to run the mechanism since it needs to know when each sample was used. Extended replacement then assigns a relevance attribute to each sample and replaces the ones with lowest values of this new attribute. Nevertheless, this mechanism is subject to a problem similar to LFU, since a frequently used sample which becomes unrepresentative will not be easily replaced.

Usage Control keeps only those detectors more frequently and recently used. It can overcome some of the issues of the previous algorithms based on usage of samples as discussed in [67]. For example, even if a detector/sample is used too many times and becomes unrepresentative, it could be quickly replaced if it is not used for a while. Hence, even if the frequency of usage of a detector/sample is the highest among all samples, it can be replaced if it has not been used recently. Usage Control 2 [66] implements another interesting proposal: a gallery of variable size. Some versions of Usage Control does not always replace a sample [67]. If it considers that the current biometric reference is representative, the replacement does not occur, as described in the criterion Usage of detectors/samples described in Section 3.2.

A recent work employed transfer learning to adapt SVM models [13]. The proposal obtained good results. However, the evaluation methodology described in these works mentions that the samples used for adaptation are labeled. Nevertheless, in a practical scenario, the true labels may not be available. It is still unclear whether it can obtain good performance if predicted (and not true) labels are used for adaptation. Table 4 summarizes the discussion of the replacement mechanisms.

3.5.3 Multi-gallery mechanisms. A multi-gallery mechanism manages two or more galleries/models to perform adaptation and can apply different adaptation mechanisms to each one. This can be interesting to combine the benefits of different adaptation mechanisms into a single one. Some implementations are presented next.

• *Double parallel* (Procedure 3 of the supplementary material). It consists of using two galleries, where one is adapted by Growing window and another is adapted by Sliding window [29]. The classification and adaptation then consider the average of the scores obtained by both galleries. An incremental version [66] allows using the growing window without the unlimited memory issue when using the classification algorithm of [47].

• *Co-Update* (Procedure 2 of the supplementary material). It is an implementation of the concepts from Co-training [10] to adaptive biometric systems [87, 97]. This paper considers the implementation described in [91]. Co-Update is applied to a multi-modality scenario using two galleries, each one for a different biometric modality (*e.g.*, one for face and another for fingerprint). It assumes that two biometric samples (one for each modality) are provided for each query. If the classifier

Table 4. Comparison of replacement adaptation mechanisms.

Mechanism	Advantages	Drawbacks
Sliding window [41]	- Simple, just replaces the oldest samples considered less representative.	- Oldest samples may be more representative.
MDIST and DEND [20]	- MDIST can exploit common representative characteristics, while DEND is able to represent larger intra-class variability.	- Both can be computationally intensive.
Least frequently used (LFU) [101]	- Replaces less frequently used patterns.	- May not replace a frequent used sample that becomes unrepresentative.
Least recently used (LRU) [101]	- Replaces less recently used patterns.	- May be expensive, since it requires to store when each sample is used.
Extended replacement [101]	- Assigns a relevance attribute to each sample, which can be used to replace less representative samples.	- Problem similar to LFU, since it may not replace a frequently used sample that becomes unrepresentative.
Usage Control [65–67]	- Does not change the biometric reference if all patterns are being used, which could mean that the user characteristics have not changed.	- May not properly adapt the reference if all patterns were recently used and the user starts to change its characteristics.
Transfer learning [13, 121]	- Can adapt SVM models without the need to retrain it.	- Uses labeled samples to adapt the SVM model.

trained for modality A confidently classifies the corresponding query, the one from modality B is added to the corresponding gallery. The opposite also applies, if the classifier trained for modality B confidently classifies the query from its modality, the query for modality A is added to the gallery of modality A. Co-Update is similar to the cross-training mechanism presented in [74, 75]. Another application of Co-training to adaptive biometric systems was presented in [117], where a single modality was considered (face recognition). In their work, each of the two classifiers considered a different view of the face image.

Poh *et al.* [74] also discussed the application of Co-training to adaptive biometric systems. It was studied a system where there was one gallery for face recognition and another for speech recognition. Taking advantage of the availability of two modalities, logistic regression combined face and speech scores to obtain the final fused score which was then used to infer the samples for adaptation. The proposed strategy was named fusion-based co-training.

• *Enhanced template update (ETU).* Adaptation mechanisms are generally only interested in the queries classified as genuine. Recently, it was proposed to make use of all queries, including those classified as impostor [63]. In order to implement it, ETU manages two galleries: one for queries classified as genuine and another for queries classified as impostor. The ETU framework then employs both galleries to support classification and adaptation.

• *Ensembles.* El Gayar *et al.* [17] proposed to use several classifiers in an ensemble configuration to address the problem of having a limited amount of labeled enrollment samples. Another work which also applied ensembles for adaptive biometric systems is [68], where different adaptation mechanisms were combined in an ensemble. This work was later extended in [64], where a proposal to adapt the meta-classifier was presented.

One of the first multi-gallery mechanisms proposed in the literature, Co-Update [87, 91, 97], is applied to multi-modal systems. This adaptation mechanism can adapt the biometric reference to larger changes due to the use of two biometric modalities. For example, in case of an abrupt change in one biometric modality, while the other one does not change, the biometric system would be able to capture this large change and adapt the reference. Otherwise, an adaptation mechanism

Table 5. Comparison of multi-gallery adaptation mechanisms.

Mechanism	Advantages	Drawbacks
Co-Update [87, 91, 97]	- Can adapt the reference even for large intra-class variation.	- Requires two biometric modalities working in parallel with aging patterns not correlated.
Double parallel [29]	- Can combine two adaptation strategies, one preserving initial patterns (Growing) and another maintaining only the latest patterns (Sliding).	- Can increase the amount of used memory indefinitely, although a solution for a specific classification algorithm has been presented in [66].
Enhanced template update [63]	- Manages a genuine and an impostor gallery, making use of all received queries to adapt them.	- Classification errors can result in unreliable information on both galleries.
Ensembles [68]	- Increased classification reliability by the use of ensembles.	- Needs more processing time than a single classifier system due to the use of several of them in an ensemble configuration.

that uses just one gallery would not be able to decide whether this abrupt change is an impostor attempt or not.

Later, Double parallel [29] was proposed. It manages two galleries for a single modality, each adapted by a different adaptation mechanism. One gallery uses Growing, thus preserving the initial user patterns, while the other gallery uses Sliding, thus maintaining only the most recently used patterns. As a result, Double parallel can combine the models obtained from both galleries to support classification and adaptation. Since Double Parallel uses Growing, one of its galleries can increase without any limit over time. In [66], the authors proposed an incremental solution to deal with this problem for the classification algorithm of [47].

Most adaptation mechanisms only consider galleries for genuine data and, consequently, they discard queries classified as impostor. Enhanced template update (ETU) [63], conversely, manages a genuine and an impostor gallery. Hence, all queries, even those classified as impostor, are used for adaptation. ETU then combines both galleries to support classification and adaptation.

Ensembles of classifiers have also been used in the literature of adaptive biometric systems [17, 64, 68]. Although the use of additional classifiers can result in higher use of computer resources, the robustness of the classification and adaptation can be increased. The fusion-based co-training proposed by Poh et al. [74] can be considered an example of this approach as well, where a classifier is associated with a biometric modality and both results are fused.

A summary of the discussion on multi-galley mechanisms is presented in Table 5.

3.5.4 Selection mechanisms. Selection mechanisms, also known as template selection [21], are used to select representative samples/templates for the user. These mechanisms can be used to reduce the size of the user gallery after adaptation [111]. Some implementations are presented next.

• *Selection based on clustering* [111]. It is based on the algorithms used for replacement discussed in Section 3.5.2. DEND applies a hierarchical clustering algorithm, which outputs a dendrogram on which a pre-defined number of clusters is identified. For each cluster, the medoid element (sample) is kept in the user gallery, while the other samples are discarded. The other mechanism, MDIST, sorts the samples by their average distance to all other samples. Those samples with the lowest average distance are kept in the user gallery, while the others are discarded. For both mechanisms, the number of samples to be kept needs to be defined. This number should be lower than the amount of samples in the gallery.

Table 6. Comparison of selection mechanisms.

Mechanism	Advantages	Drawbacks
Selection based on clustering	- Can reduce the size of the gallery using clustering algorithms.	- Can be computationally intensive for large galleries.
Selection based on editing	- Can reduce the size of the gallery using NN-based algorithms.	- When strong gallery size limitations are imposed, the output gallery can be negatively impacted. - Can be computationally intensive for large galleries.

• *Selection based on editing*. In [21], the authors proposed the use of algorithms based on the nearest neighbor algorithm to select the most representative samples for a user: Condensed NN (CNN) [33], Selective NN (SNN) [95], Reduced NN (RNN) [22] and Edited NN (ENN) [114].

Both methods, selection based on clustering [111] and based on editing [21], can be used to reduce the gallery size after adaptation. This can be particularly important for additive mechanisms, such as Self-Update [90, 98]. Freni *et al.* [21] compared both types of mechanisms and showed that editing mechanisms can obtain better performance than clustering mechanisms. A summary of this discussion is shown in Table 6.

4 EVALUATION OF ADAPTIVE BIOMETRIC SYSTEMS

The evaluation of adaptive biometric systems differs from the evaluation of standard non-adaptive biometric systems. First, it involves dealing with an additional process: the adaptation. Second, usually, the sequence of queries is chronologically ordered to assess the adaptation of the biometric reference over time. Third, this implies that the datasets used for such an evaluation are expected to contain several samples per user captured over time. This section discusses several aspects related to the evaluation of adaptive biometric systems, including datasets, metrics and methodologies.

4.1 Modalities and datasets

The availability of suitable datasets for the evaluation of adaptive biometric systems is limited. A possible reason is the intrinsic difficulty to acquire data for such kind of study as these datasets need to contain several samples per user. Ideally, they should be obtained at different acquisition sessions, either with distinct acquisition conditions or separated by a certain amount of time, to justify the use of an adaptive biometric system. Table 7 lists some datasets in the literature.

The analysis of the existing literature shows that the number of users and the time period/sessions significantly differ among the datasets. While it is generally true that a higher number of users and longer sessions can result in a more reliable estimate of performance, the variability in the nature and context of previous experiments means that it is extremely hard to compare different adaptation techniques. It also shows a higher number of datasets for physical biometric modalities, mainly for face and fingerprint. This fact may explain the higher number of studies for physical modalities in the field. Since the rate of change in physical biometric modalities is likely to be smaller than the behavioral ones, it is difficult to extrapolate the findings to these modalities.

4.2 Metrics

Most studies evaluate adaptive biometric systems with the same metrics as those used to evaluate standard, non-adaptive biometric systems. In this section, we shall first discuss the standard metrics and then highlight a few others that are more specific to the evaluation of adaptive biometric systems. These metrics can be applied to evaluate a biometric system in an experiment, for instance.

Table 7. Modalities and datasets used in the evaluation of adaptive biometric systems. For accelerometer biometrics datasets, the number of users and samples were obtained after the procedure described in [69].

Modalities	Datasets	# Users	Period/Sessions
Keystroke dynamics	GREYC [26]	100	2 months (5 sessions)
Keystroke dynamics	GREYC-Web [27]	118	more than 1 year
Keystroke dynamics	CMU [43]	51	8 sessions
Face	AR Face Database [50]	116	14 days (2 sessions)
Face	Dataset from [6]	40	2 sessions
Face	BANCA 2D [3]	52	12 sessions
Face and fingerprint	DIEE multi-modal [49]	49	1.5 years (10 sessions)
Fingerprint	FVC-2002 DB2 [48]	110 fingers	3 sessions
Fingerprint	Dataset from [111]	50 fingers	aprox. 4 months (2 sessions)
Iris	Fenker [19]	322	aprox. 4 years
Voice	ELDASR [107]	50	20 samples per user
Accelerometer biometrics	WISDM 1.1 [44]	33	180.55 samples per user (average)
Accelerometer biometrics	WISDM 2.0 [45]	131	213.34 samples per user (average)
Ocular images	VISOB [83]	550	2 visits (2 sessions/visit)
Fingerprint	LiveDet2011 [115]	200	6 materials

Monitoring the error rates could improve the recognition performance of a biometric system as discussed in [106]. However, controlling adaptation strategies by monitoring its performance over time is still an open issue.

4.2.1 Common metrics from biometrics. As mentioned earlier, several metrics used to assess the recognition performance of adaptive biometric systems are common to the evaluation of other biometric systems. They are detailed below [34, 74, 79]:

- *FNMR (False Non-match Rate)*: the rate of genuine *attempts* that were wrongly classified as impostor. The FNMR for a given user j in verification mode is computed as shown in Equation (5), where Q_j^G is the set of true genuine queries compared to the biometric reference of the genuine user j. As seen in the equation, the $FNMR_j$ depends on the parameters adopted for the verification θ_j^{verify}. In the case of an adaptive biometric system, the biometric reference ref_j can change over time and, therefore, the $FNMR_j$ also depends on the set of parameters for the adaptation process θ_j^{adapt} and how the adaptation set $\mathcal{A}$ is formed. In Equation (5), t is the time, therefore, the query $\mathbf{q}_t$ is matched against the biometric reference ref_{j_t} at the time t. In a non-adaptive biometric system, ref_{j_t} is the same regardless of the time t, since the biometric reference is not modified.

$$FNMR_j(\theta_j^{verify}) = \frac{|\ \{\mathbf{q}_t \mid \mathbf{q}_t \in Q_j^G \wedge impostor = testVerify(ref_{j_t}, \mathbf{q}_t \mid \theta_j^{verify})\}\ |}{|\ Q_j^G\ |} \tag{5}$$

In order to report the global FNMR, the average from all users can be computed, as shown in Equation (6). Another approach computes the metric considering the number of genuine queries from all users at the same time as shown in Equation (7). Note that when the number of genuine queries is different among the users (*e.g.*, GREYC-Web dataset [27]), these two methods to compute the global FNMR can result in different values. The first one

gives the same weight to each user (Equation (6)), while the second one gives more weight to those users which contain a higher number of genuine queries (Equation (7)).

$$FNMR(\theta^{verify}) = \frac{\sum_{j \in \mathcal{J}} FNMR_j(\theta_j^{verify})}{\mid \mathcal{J} \mid} \tag{6}$$

$$FNMR(\theta^{verify}) = \frac{\sum_{j \in \mathcal{J}} \mid \{\mathbf{q}_t \mid \mathbf{q}_t \in Q_j^G \wedge impostor = testVerify(ref_{j_t}, \mathbf{q}_t \mid \theta_j^{verify})\} \mid}{\sum_{j \in \mathcal{J}} \mid Q_j^G \mid} \tag{7}$$

The FRR (False Rejection Rate) is a related metric that considers the FTA (Failure to Acquire Rate) (Equation (8)). FTA measures the rate in which a biometric system fails to obtain a biometric sample.

$$FRR(\theta^{verify}) = FTA + FNMR(\theta^{verify}) \times (1 - FTA) \tag{8}$$

- *FMR (False Match Rate)*: rate of impostor *attempts* that were wrongly classified as genuine. The FMR for a given user j in verification mode is computed as shown in Equation (9), where Q_j^I is the set of true impostor queries compared to the biometric reference of the genuine user j. As seen in the equation, the FMR_j depends on the parameters adopted for the verification θ_j^{verify}. In the case of an adaptive biometric system, the biometric reference ref_j can change over time and, therefore, the FMR_j also depends on the set of parameters for the adaptation process θ_j^{adapt} and how the adaptation set $\mathcal{A}$ is formed. In Equation (9), t is the time, therefore, the query $\mathbf{q}_t$ is matched against the biometric reference ref_{j_t} at the time t. In a non-adaptive biometric system, ref_{j_t} is the same regardless of the time t, since the biometric reference is not modified.

$$FMR_j(\theta_j^{verify}) = \frac{\mid \{\mathbf{q}_t \mid \mathbf{q}_t \in Q_j^I \wedge genuine = testVerify(ref_{j_t}, \mathbf{q}_t \mid \theta_j^{verify})\} \mid}{\mid Q_j^I \mid} \tag{9}$$

The global FMR, which measures the average FMR from all users, is shown in Equation (10). Another approach is to simply compute the metric considering the number of impostor queries from all users at the same time as shown in Equation (11). Note that when the number of impostor queries is different among the users, these two methods to compute the global FMR can result in different values. In some evaluation methodologies, the number of impostor queries is a function of the number of genuine queries [63]. In the first method, each user has the same weight (Equation (10)). In the second method, users which contain a higher number of impostor queries receive a higher weight (Equation (11)).

$$FMR(\theta^{verify}) = \frac{\sum_{j \in \mathcal{J}} FMR_j(\theta_j^{verify})}{\mid \mathcal{J} \mid} \tag{10}$$

$$FMR(\theta^{verify}) = \frac{\sum_{j \in \mathcal{J}} \mid \{\mathbf{q}_t \mid \mathbf{q}_t \in Q_j^I \wedge genuine = testVerify(ref_{j_t}, \mathbf{q}_t \mid \theta_j^{verify})\} \mid}{\sum_{j \in \mathcal{J}} \mid Q_j^I \mid} \tag{11}$$

A related metric is FAR (False Acceptance Rate), which has almost the same definition of FMR, similarly to the case of FNMR/FRR. FAR also considers the FTA, as shown in Equation (12).

$$FAR(\theta^{verify}) = FMR(\theta^{verify}) \times (1 - FTA) \tag{12}$$

- *HTER (Half Total Error) and balanced accuracy*: HTER is defined by Equation (13) as the average between FNMR and FMR. This metric combines the results from both FNMR and FMR into a single value, making the performance evaluation simpler. This measure can also be defined using the balanced accuracy *BAcc* [51], defined in Equation (14).

$$HTER(\theta^{verify}) = \frac{FNMR(\theta^{verify}) + FMR(\theta^{verify})}{2} \tag{13}$$

$$BAcc(\theta^{verify}) = 1 - HTER(\theta^{verify}) \tag{14}$$

- *EER (Equal Error Rate)*: it is the value when FNMR is equal to FMR. This metric can be seen as a particular case of HTER, when $FMR = FNMR$.

Several studies only report EER results, leading to some drawbacks, as mentioned in [4]. The use of this metric requires testing different parameter values (*e.g.*, decision threshold) on the test data, until false match equals false non-match. This procedure to obtain the threshold may not be feasible in a practical scenario. Moreover, if this measure is computed over time, the parameter values that result in the EER may change over the sessions [24]. In view of these problems, a better approach to assess the recognition performance is to tune the parameters in the enrollment data and then apply the obtained parameters to the test data. A consequence of this procedure is that the EER cannot be computed, but, instead, the couple of FMR and FNMR for a given set of parameter values is reported.

These metrics can also be obtained over time. The study in [89] claims to be the first one in the area to compute results over time, instead of just reporting it globally. Later, a plot to report performance metrics over time in the context of a biometric data stream was proposed [63]. Overall, this plot extracts the metric by using a sliding window over the biometric data stream. The average values for all users at each window are plotted. In some datasets, the number of samples varies among the users and, therefore, the later parts of the plot would be the average of a lower number of users (the ones which contain the higher number of samples only). In order to deal with this problem, the plot also shows the interval based on the *standard error of the mean* (shaded area) as in Equation (15) (CI_i: confidence interval), which makes use of *SE* (Standard Error) calculated in Equation (16). The notation std_i denotes the standard deviation among the values at window i and $users_i$ denotes the number of users with available data at window i. This interval provides additional data to support the discussion of the results.

$$CI_i = mean(measure)_i \pm 1.96 \times SE_i \tag{15}$$

$$SE_i = std_i / \sqrt{users_i} \tag{16}$$

Figure 6 illustrates this plot, where the FNMR over time comparing non-adaptive and adaptive biometric systems is shown. *Self-Detector* and *M2005* represent non-adaptive biometric systems. The versions on the right (*Sliding*, *Growing*, *DB*, *IDB*) represent the adaptive biometric systems over the same classification algorithm. Note that the adaptive biometric systems managed to reduce the FNMR, which is a good result. In addition, the standard deviation among the users is higher when adaptation is not applied in this dataset.

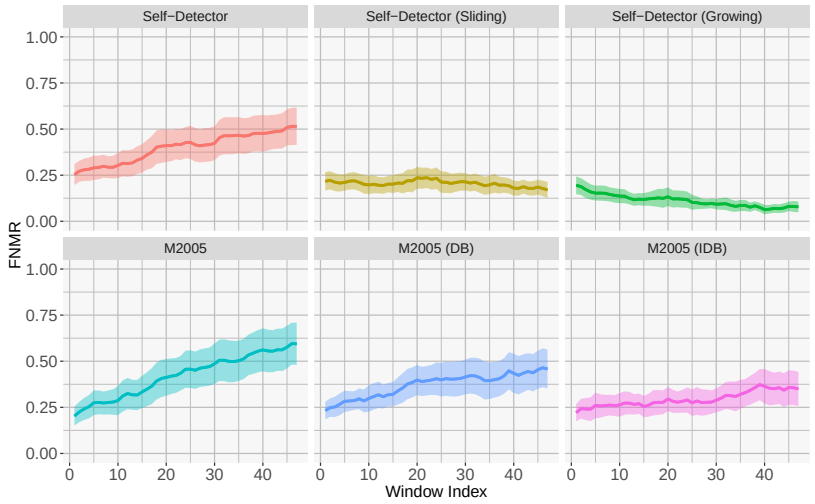


Fig. 6. FNMR over time comparing non-adaptive and adaptive biometric systems (CMU dataset - keystroke dynamics). *Self-Detector* and *M2005* represent non-adaptive biometric systems. The versions on the right (*Growing*, *Sliding*, *DB*, *IDB*) represent the adaptation strategies.

4.2.2 Metrics specific to adaptive biometric systems. In addition to the previous metrics, some metrics specific to adaptive biometric systems have been proposed. As shown in Section 3, most adaptation strategies only use queries classified as genuine from the set $\mathcal{A}$ to adapt the biometric reference (Equation (4)). Based on this concept, two metrics were proposed to assess the correct use of genuine samples by the adaptation strategy [29]:

- *IUSR (Impostor Update Selection Rate)*: rate of impostor samples involved in the adaptation process, as defined by Equation (17).

$$IUSR = \frac{number\ of\ impostor\ samples\ involved\ in\ the\ adaptation\ process}{number\ of\ tested\ impostor\ samples} \tag{17}$$

- *GUMR (Genuine Update Miss Rate)*: rate of genuine samples not involved in the adaptation process, as defined by Equation (18).

$$GUMR = \frac{number\ of\ genuine\ samples\ not\ involved\ in\ the\ adaptation\ process}{number\ of\ tested\ genuine\ samples} \tag{18}$$

4.3 Evaluation methodology

Unfortunately, there is no standard methodology to evaluate adaptive biometric systems in the literature [30]. Several methodologies that differ in several aspects have been adopted. This section discusses them.

4.3.1 Impostor samples in the adaptation process. Recent studies dealing with adaptation considered that the set of biometric samples for adaptation $\mathcal{A}$ (Equation (4)) was a set of samples without the true label. Thus, only labels obtained from the classification algorithm were available, so they were subject to wrong prediction from the classifier. This better simulates a practical scenario

where true labels are usually not available. Consequently, the set $\mathcal{A}$ may contain impostor samples resulted from misclassification.

However, early investigation on adaptive biometric systems did not consider the possibility of impostor attack during the adaptation. This is observed in [111], which employed a dataset with two sessions (100 samples per session). Each session was divided into two parts, one for training/enrollment and another for test. Hence, session 1 was divided into TRAIN1 and TEST1, while session 2 was divided into TRAIN2 and TEST2. In that work, the evaluation was performed as follows: TRAIN1 was used to obtain the initial biometric reference, which was later tested on samples from TEST1. Samples from other users were regarded as impostors during the test. Afterward, the biometric reference was adapted using TRAIN2. Later, the adapted biometric reference was tested on TEST2. Note that by doing this, it is assumed that a set of true genuine samples is provided for adaptation.

According to [77, 91], impostor attacks during the adaptation process were not considered in [97, 98]. Another study that did not consider impostor samples in the set $\mathcal{A}$ is [41]. As mentioned in [24, 28], the experiments on [41] only employed true genuine samples for adaptation. In [41], each user was enrolled using 10 samples and, for test, there were 75 genuine samples plus 75 impostor samples. Although not entirely clear, the graphs from Figure 4 of that paper indicate that a separate set of genuine samples was used for adaptation.

4.3.2 Ratio of impostor samples. A related aspect is the ratio of impostor samples that can be part of the adaptation set $\mathcal{A}$. A high ratio can result in several errors during the adaptation process. In [91], the adaptation set $\mathcal{A}$ contained 10 genuine samples and 5 random impostor samples, so the ratio of impostors was 33.3%. Another study [29] adopted the ratio of 30% of impostor samples. It assumed a scenario where the genuine user was the most frequent user of the biometric system, which is a valid assumption in several cases.

Later, different ratios of impostor samples were investigated in [31] where the samples from the first session were used for enrollment. Then, the samples from the remaining sessions were used for test and adaptation using *pools*. A *pool* was defined as a sequence of query samples, containing both genuine and impostor. The ratio of impostors in the *pools* ranged from 30% to 80%. One pool was generated for each session in the dataset. By doing this, the performance metrics could be assessed over time, one for each session.

The same ratio of impostors of 30% was also adopted in [63, 66, 70]. However, in those studies, a distinct method was adopted to select the impostor samples. The evaluation methodology adopted there, named *user cross-validation for biometric data streams*, divided the list of user indexes using cross-validation, so k folds were obtained (each fold was a disjoint subset of the user indexes). One fold was regarded as the unregistered set of users and the remaining folds formed the registered set of users $\mathcal{J}$. The experiments were executed for all k combinations of folds, so all users were considered once as an unregistered user. Among the 30% of impostor samples, there was a 50% probability of obtaining an impostor sample from the unregistered set (external attack simulation) and a 50% probability of obtaining a sample from another user $i \neq j$ (internal attack simulation) as impostor.

4.3.3 Adaptation to time vs condition. Template aging is one of the main motivations for adapting a biometric reference. This is clear in keystroke dynamics in which the typing rhythm changes over time. However, the biometric reference may need adaptation to deal with different acquisition conditions too, which is not necessarily due to aging. For instance, in face recognition, if the enrollment uses samples of just one pose (*e.g.*, frontal), the system would need to adapt the reference later to include variations in the pose of the same user.

The methodology described in the last section from [31] is one that mainly dealt with adaptation due to aging. This is because the first session was used for training and next ones were left for test and adaptation, following the chronological order.

Conversely, the experiments in [74] is an example of methodology which mainly dealt with adaptation to different conditions. That work used a dataset which contained data under three different conditions: controlled (sessions 1-4), degraded (sessions 5-8) and adverse (sessions 9-12) [3]. Session 1 was used for enrollment, then sessions 2 to 4 for test. Next, session 5 was used for adaptation and sessions 6 to 8 for test. Finally, session 9 was used for adaptation and sessions 10 to 12 for test. Impostor samples were included in the adaptation sets too.

Adaptive biometric systems can be used to adapt the biometric reference to changes either due to time or due to different capture conditions. Some studies on physical biometric modalities seem to mainly deal with changing conditions instead of changes uniquely due to time, which is the case of that study.

4.3.4 Poisoning attacks to adaptation. Poisoning attacks in adaptive biometric systems consist in progressively introducing impostor samples in the adaptation process, in a way that the biometric reference is modified until it can better recognize an impostor. As a result, it may also not be able to recognize the actual genuine user anymore. These attacks are not simulated in most evaluation methodologies for adaptive biometric systems.

The work that claims to be the first to raise such an issue in the area is [6]. In order to evaluate this attack, the authors used a dataset for face recognition containing 60 samples per user. A random subset of 10 samples was used for the enrollment and another subset of 10 samples was used for parameter tuning. The remaining 40 samples were then used for the test. Then, a separate set of poisoning samples was used to adapt the biometric reference. Nevertheless, their work only considered that the biometric reference is adapted with impostor patterns from the generated poisoning set. This may not correspond to a practical scenario, since both genuine and impostor samples can be used for adaptation and, consequently, the negative effect of poisoning could be reduced.

4.3.5 Separate and joint sets for test/adaptation. Most previous evaluation methodologies can be divided into two groups: *separate sets* or *joint set* for test and adaptation. A previous review in the area adopted this criterion to classify performance assessment approaches [78]. In the *separate sets* approach, the adaptation and test sets are disjoint and, consequently, samples used for adaptation are not part of the test. This approach assumes that the biometric system can stay a period only adapting the biometric references (without performing test/recognition). Later, the adapted biometric reference is fixed to perform recognition only. Some recent studies have also adopted the *separate sets* approach [6, 74]. However, this approach may not be the best choice in some cases since it does not make optimal usage of the available data. This is due to the non-overlapped adaptation and test sets. The optimal usage of the dataset is a critical issue in the area, particularly in view of the limited number of large datasets for studying adaptive biometric systems.

The *joint sets* for test and adaptation approach, on the other hand, share data for test and adaptation, so both sets are not disjoint. This approach also better represents a practical scenario, where the system, once deployed, has to perform the recognition of all query samples and use this data for adaptation. Hence, the system does not stop the recognition for a period of adaptation.

An important work in the area which proposed an evaluation methodology following the *joint sets* approach is [91]. A similar methodology was used in another work from the same authors in [90]. Their methodology was based on the DIEE dataset, which has several sessions per user, each containing 10 samples. The following steps are performed:

- Part A (enrollment): the first 2 samples of the first session ($t = 1$) are used for enrollment.
- Part B (adaptation): an adaptation set $\mathcal{A}$ is built for each user with the samples from current session t plus five random impostor samples. The first session used for adaptation is $t = 1$, however, in this particular case, the first two samples are discarded since they were already used for enrollment, while, in the other sessions, all 10 samples are part of the adaptation set. This adaptation set is then presented to the adaptation strategy to perform adaptation.
- Part C (test): the adapted biometric reference is used to test on the next session. The first test session is $t + 1$. Biometric samples from the same session from all other users are regarded as impostors to compute the performance metrics. Note that the biometric reference is not adapted during the test. When the test is finished on session $t + 1$, Part B is launched again, though on session $t + 1$ this time. The adapted biometric reference is then tested on session $t + 2$ and so on.

Note that, in this methodology, the last session is used only for test and the very first session is only part of the enrollment and adaptation. However, all other sessions are used for both processes, meaning that it mainly adopts the *joint sets* for test and adaptation approach. As a result, the number of samples used for both adaptation and test is increased.

Another evaluation methodology that follows the *joint sets* approach is [29] and its modification to include variable impostor ratios too [31]. As previously described, a *pool* is generated for each session and they are applied for test and adaptation, so the same data is used by both processes. The work from [63, 66] also adopted this approach as the same biometric data stream used for test is also the input for adaptation.

4.3.6 Online vs Offline adaptation. As discussed in Section 3, the periodicity of adaptation can change. There are two general categories: offline and online adaptation. In the offline adaptation, the biometric reference keeps unchanged for some time, then it is adapted at specific periods. Conversely, in the online adaptation, the biometric reference is adapted after each query is presented to the biometric system.

The methodology from [91] described earlier in this paper is an example of offline adaptation, since the biometric reference keeps unchanged during the test, while the methodology adopted in [31] deals with online adaptation. The *pool* is presented query by query to the biometric system, which will perform recognition and then adapt the biometric reference.

4.3.7 Chronological order. Usually, the evaluation of adaptive biometric systems respects the chronological order of the biometric samples. The enrollment should be done using the oldest samples and the test using the newest samples, in chronological order, to properly evaluate how the biometric system adapts the biometric reference to changes over time. As a result, the biometric reference is adapted to progressive changes observed over time. Modifying the order of the samples during the test can change how the biometric reference is adapted and, therefore, if the goal is to study changes due to time, the obtained results would be unreliable.

In [31], for instance, the samples from the *pool* are randomly interleaved (between genuine and impostor samples), but the chronological order of the genuine samples is maintained. Another work that respects the chronological order is [57], but it is not the case for all studies in the area, such as [6], which studied the effect of poisoning attacks. In that study, random samples were used for enrollment, so test samples may be newer than the enrollment ones.

4.3.8 Division into sessions and biometric data streams. As discussed in the previous section, several evaluation methodologies used the division into sessions to guide the assessment of the biometric systems. In [90, 91], for example, the session division information is used to guide when the adaptation process is launched. In other studies, such as [29, 31], the session division information

is used to guide the generation of the *pools*, as one *pool* is obtained from each session. The decision threshold may also change over the sessions since the results are reported in terms of EER.

The information regarding the session division may not be available in a practical scenario. In light of this fact, in the studies from [62, 63, 66, 70] which used the *user cross-validation for biometric data streams* methodology, a biometric data stream is generated for each user ignoring the session division. It works by joining all the sessions into a single one, then the first samples are used for enrollment (where parameter tuning is performed too) and the remaining ones are used to form a biometric data stream. This biometric data stream is a sequence of queries presented, sample by sample, to the biometric system, which will return the $label^p$ for each query. The decision to adapt or not the biometric reference in the meantime is up to the adaptation strategy. As a result, the decision to adapt a biometric sample is taken by the adaptive biometric system without the help of additional information, such as the session division.

5 OPEN CHALLENGES

After having reviewed previous research, this section presents some challenges and opportunities in the area of adaptive biometric systems.

5.1 Large scale adaptive biometric systems

So far, experimental studies on adaptive biometric systems have been evaluated in small to medium datasets ranging from 33 to 550 users (see Table 7). Although there are real application scenarios that involve a similar number of users, there are also applications whose number of users can be much higher (*e.g.*, border control). Thus, the application of adaptive biometric systems in a large scale scenario is a current challenge. Research in this scenario may add new questions, like the impact on time and space imposed by the adaptation strategies. The findings will certainly find its way in commercial applications, as discussed in Section 5.6. Suitable, preferably public, datasets for studying large scale adaptive biometric systems are also necessary (see Section 5.2 and 5.3).

5.2 Acquisition of datasets suitable to evaluate adaptive biometric systems

Datasets should meet some important requirements to be used to evaluate adaptive biometric systems (see Section 4.1). These datasets need to contain several samples per user and, ideally, they should be acquired at different acquisition sessions. Currently, there are some public datasets, however, additional ones with a larger number of users and sessions are still needed. They would allow, for example, the evaluation of large scale adaptive systems, discussed in the previous item. The acquisition of these datasets can demand a large effort, since it requires to obtain biometric data for the same users during long periods. These databases could also be useful to implement score normalization [76] for adaptive biometric systems, which can benefit research in the area.

5.3 Generation of synthetic datasets for template update evaluation

While it is very difficult to collect biometric data to evaluate standard biometric systems, it is even more difficult to collect temporal biometric data to evaluate adaptive biometric systems. An alternative to deal with this problem is to generate synthetic data to be able to evaluate proposed algorithms on large scale datasets. This approach has been used for digital fingerprints with the definition of the SFINGE software [11] that has demonstrated its interest in the Fingerprint Verification Competition (FVC). Recent studies on other modalities, such as keystroke dynamics, open new perspectives for the generation of synthetic datasets [58] that could fit the requirements for adaptive biometric systems. Nevertheless, an open question that remains is how to artificially simulate realistic changes over time. It is known that changes may occur in different fashions

depending on the biometric modality [30]. Therefore, the simulation should take into account such differences. Future work may also model how changes impact different biometric modalities.

5.4 Attacks to the biometric system

Biometric references can be poisoned when impostor samples are used to adapt genuine references. This usually occurs by exploiting the adaptation process to progressively introduce impostor patterns into genuine biometric references. Consequently, an impostor can, not only impersonate, but also deny access to the genuine user. This is related to adversarial machine learning, which studies vulnerabilities of learning algorithms and the respective countermeasures [7]. A few papers that deal with this issue for adaptive biometric systems are [5, 6, 113]. Since adaptation strategies usually rely on how well the classification algorithm performs adaptation, classification errors can result in the inclusion of impostor patterns into genuine biometric references.

There are some categories of attacks to biometric system [7], such as spoofing, replay, hill climbing, malware infection and reference database attacks. Spoofing is the presentation of a fake biometric trait to a biometric sensor (*e.g.*, by using a silicon model of a finger). Replay involves stealing a true raw biometric data (*e.g.*, fingerprint image). This stolen data can be introduced in the biometric system. This data can also be modified before the introduction, thus harmfully adapting the true biometric reference. Hill climbing attacks simulate and introduce biometric data into the biometric system in order to gradually obtain biometric data similar to that of the target user. Malware infections affect the whole system integrity and database attacks either steal or introduce damaging information by a direct attack on the system database [6].

Some countermeasures have been proposed to deal with the previous problems. For instance, liveness detection systems [23, 93] can avoid spoofing attacks. Several countermeasures can also be combined [92]. Similar approaches could provide effective solutions to protect against various attacks. Moreover, another promising approach is the proposal of measures specifically tailored for the evaluation of how different types of attacks affect the performance of adaptive biometric systems. Current research in adaptive biometric systems usually considers only the well-known zero-effort attacks, which use samples that do not belong to the target user to simulate attacks. Although this is a valid approach, those samples are from users that did not deliberately studied the target user to be attacked.

5.5 Adaptive biometric system able to adapt to both condition and time

Several adaptation strategies have been proposed over the last years, as presented throughout this paper. However, some of them may be more suitable to adapt biometric references to changes due to time (*e.g.*, *Sliding* and *Usage Control*), while others may be more suitable to adapt biometric references to changes due to new acquisition conditions (*e.g.*, *Self-Update*). This is also reflected in the evaluation methodologies. As discussed in Section 4.3.3, some methodologies adopted in previous studies mainly evaluate aging, while others mainly evaluate new conditions.

In view of this scenario, future work could evaluate adaptive biometric systems to assess how they adapt to both aging and changing acquisition conditions. This may require new methods to evaluate these adaptation cases. In addition, adaptive strategies specifically designed to deal with both aging and condition changes could be proposed. A formalization for such has been presented by Poh *et al.* [77].

5.6 Adaptive biometric sytems for commercial applications

Despite all the benefits of using adaptive biometric systems discussed in this paper, to the best of our knowledge, there is very few explicit indication on the use of adaptive biometric strategies in

commercial applications; for example, *Apple's Face ID* [36] states that an adaptation is performed with the augmentation of the face data.

The adoption of adaptive biometric systems in commercial applications can be considered a current challenge, especially for applications that focus on daily-life as they cannot be as much controlled as border control-like applications. A key aspect that may have prevented their use is the vulnerability to impostor attacks [77] (see Section 5.4). A mechanism to prevent the inclusion of impostor patterns in the genuine reference might encourage vendors to include adaptation strategies on their products.

5.7 Adaptive biometric strategies in mobile applications

Biometric recognition solutions are nowadays widely used in commercial systems for securing physical and logical access, as discussed in Section 1. Several of these systems are used in mobile applications, highlighting the importance of considering mobile devices. Major challenges for real-time biometric recognition on mobile phones are computing performance, memory storage and security. Therefore, these challenges are also present for adaptive strategies on mobile devices. As presented throughout the paper, many adaptation strategies already meet this requirement, since they use simple update procedures.

Nonetheless, even for the adaptation strategies which may require high computation power, it is still feasible to adopt them by moving computing tasks to servers in the cloud. A biometric system under this setting would also require a secure communication channel to perform the authentication and adaptation processes.

An example of study which deals with mobile devices is [82], which applied co-training in this scenario. Nevertheless, there is still a limited number of studies which evaluate adaptive biometric systems on mobile devices. In addition to dealing with hardware limitations, research on mobile adaptation should also acquire data from the sensors on these devices over time.

6 CONCLUSION

Biometric systems can verify the identity of an individual based on biological, morphological and behavioral characteristics. They have been successfully used in several applications. Characteristics used for recognition should meet some properties [37], as discussed at the beginning of this paper: universality, distinctiveness, permanence and collectability. However, recent studies have shown that the permanence is not met for several biometric modalities [29, 70, 77, 81, 97]. This is due to several reasons, including aging and changing conditions, as discussed in Section 2.2. In order to deal with this problem, adaptive biometric systems have been proposed as one of the solutions. This is a relatively new field of study in biometrics.

The aim of this paper has been to provide a wide review of adaptive biometric systems, covering aspects such as formalization, terminology, sources of variations over time, adaptation strategies, evaluation methodology and open challenges. To the best of our knowledge, this is the most up-to-date and complete review of adaptive biometric systems.

Thanks to the proposed taxonomy for adaptation strategies presented in this paper, the reader has a broad view of works in adaptive biometric systems and can easily compare them. Adaptation strategies were divided into modules, namely: reference modeling, adaptation criterion, adaptation mode, adaptation periodicity and adaptation mechanism.

Another contribution of this paper is discussing the distinct evaluation methodologies that have been adopted in previous work. The evaluation of adaptive biometric systems differs from the evaluation of standard non-adaptive biometric authentication systems. Common evaluation metrics have been redefined to be properly expressed in the context of adaptive systems and specific metrics

have been described as well. To standardize the evaluation protocol of adaptive biometric systems would be an important advance to the field.

A discussion about the open challenges in adaptive biometric systems is another key aspect of this paper. The discussion points out key gaps in the literature that could be explored in future research. We believe that future studies should focus on the definition of algorithms and systems suitable for commercial applications. This implies the specification of large scale adaptive systems and mobile-based systems, as well as the acquisition of natural or synthetic datasets that could provide a benchmark to evaluate these systems. Additionally, these systems must also be resistant to poisoning attacks and able to adapt to both condition and time changes.

REFERENCES

[1] Charu C. Aggarwal. 2015. *Data Classification.* Chapman and Hall/CRC.

[2] Zahid Akhtar, Arif Ahmed, Cigdem Eroglu Erdem, and Gian Luca Foresti. 2014. Biometric template update under facial aging. In *Computational Intelligence in Biometrics and Identity Management, 2014 IEEE Symposium on.* IEEE, 9–15.

[3] Enrique Bailly-Bailliére, Samy Bengio, Frédéric Bimbot, Miroslav Hamouz, Josef Kittler, Johnny Mariéthoz, Jiri Matas, Kieron Messer, Vlad Popovici, Fabienne Porée, Belen Ruiz, and Jean-Philippe Thiran. 2003. The BANCA database and evaluation protocol. In *Proceedings of the 4th international conference on Audio- and video-based biometric person authentication.* 625–638.

[4] Samy Bengio, Johnny Mariéthoz, and Mikaela Keller. 2005. The expected performance curve. In *International Conference on Machine Learning, ICML, Workshop on ROC Analysis in Machine Learning.* 1–8.

[5] B. Biggio, L. Didaci, G. Fumera, and F. Roli. 2013. Poisoning attacks to compromise face templates. In *2013 International Conference on Biometrics (ICB).* 1–7.

[6] Battista Biggio, Giorgio Fumera, Fabio Roli, and Luca Didaci. 2012. Poisoning adaptive biometric systems. In *Proceedings of the 2012 Joint IAPR international conference on Structural, Syntactic, and Statistical Pattern Recognition.* 417–425.

[7] B. Biggio, g. fumera, P. Russu, L. Didaci, and F. Roli. 2015. Adversarial Biometric Recognition : A review on biometric system security from the adversarial machine-learning perspective. *IEEE Signal Processing Magazine* 32, 5 (2015), 31–41.

[8] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning.* Springer.

[9] Avrim Blum and Shuchi Chawla. 2001. Learning from labeled and unlabeled data using graph mincuts. In *Proceedings of the Eighteenth international Conference on Machine Learning.*

[10] Avrim Blum and Tom Mitchell. 1998. Combining Labeled and Unlabeled Data with Co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory.* 92–100.

[11] Raffaele Cappelli, D Maio, and D Maltoni. 2004. SFinGe: an approach to synthetic fingerprint generation. In *International Workshop on Biometric Technologies (BT2004).* 147–154.

[12] John W Carls. 2009. *A framework for analyzing biometric template aging and renewal prediction.* ProQuest.

[13] Hayreddin Çeker and Shambhu Upadhyaya. 2016. Adaptive techniques for intra-user variability in keystroke dynamics. (2016), 1–6.

[14] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. 2006. *Semi-Supervised Learning.* The MIT Press.

[15] I. Cohen, F. G. Cozman, N. Sebe, M. C. Cirelo, and T. S. Huang. 2004. Semisupervised learning of classifiers: theory, algorithms, and their application to human-computer interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 12 (2004), 1553–1566.

[16] Luca Didaci, Gian Luca Marcialis, and Fabio Roli. 2014. Analysis of unsupervised template update in biometric recognition systems. *Pattern Recognition Letters* 37 (2014), 151–160.

[17] Neamat El Gayar, Shaban A Shaban, and Sayed Hamdy. 2006. Face recognition with semi-supervised learning and multiple classifiers. In *Proceedings of the 5th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics, Venice, Italy.* 296–301.

[18] Asma El Kissi Ghalleb, Souhir Sghaier, and Najoua Essoukri Ben Amara. 2013. Face recognition improvement using soft biometrics. In *Systems, Signals & Devices (SSD), 2013 10th International Multi-Conference on.* IEEE, 1–6.

[19] S.P. Fenker and K.W. Bowyer. 2012. Analysis of template aging in iris biometrics. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on.* 45–51.

[20] Biagio Freni, Gian Luca Marcialis, and Fabio Roli. 2008. Replacement algorithms for fingerprint template update. In *Image Analysis and Recognition.* Springer, 884–893.

[21] Biagio Freni, Gian Luca Marcialis, and Fabio Roli. 2008. Template selection by editing algorithms: A case study in face recognition. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural*

and Syntactic Pattern Recognition (SSPR). 745–754.
[22] G. W. Gate. 1972. The reduced nearest neighbor rule. *IEEE Trans. Inf. Theory,* 18, 3 (1972), 431–433.
[23] Luca Ghiani, David A Yambay, Valerio Mura, Gian Luca Marcialis, Fabio Roli, and Stephanie A Schuckers. 2016. Review of the Fingerprint Liveness Detection (LivDet) competition series: 2009 to 2015. *Image and Vision Computing* (2016).
[24] Romain Giot, Bernadette Dorizzi, and Christophe Rosenberger. 2011. Analysis of template update strategies for keystroke dynamics. In *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011 IEEE Workshop on*. 21–28.
[25] Romain Giot, Mohamad El-Abed, Baptiste Hemery, and Christophe Rosenberger. 2011. Unconstrained keystroke dynamics authentication with shared secret. *Computers & security* 30, 6 (2011), 427–445.
[26] Romain Giot, Mohamad El-Abed, and Christophe Rosenberger. 2009. GREYC Keystroke: a Benchmark for Keystroke Dynamics Biometric Systems. In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*. 419–424.
[27] Romain Giot, Mohamad El-Abed, and Christophe Rosenberger. 2012. Web-based benchmark for keystroke dynamics biometric systems: A statistical analysis. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*. IEEE, 11–15.
[28] R. Giot, C. Rosenberger, and B. Dorizzi. 2012. Can Chronological Information be Used as a Soft Biometric in Keystroke Dynamics?. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*. 7 –10.
[29] Romain Giot, Christophe Rosenberger, and Bernadette Dorizzi. 2012. Hybrid template update system for unimodal biometric systems. In *Biometrics: Theory, Applications and Systems, 2012 IEEE Fifth International Conference on*. 1–7.
[30] Romain Giot, Christophe Rosenberger, and Bernadette Dorizzi. 2012. Performance Evaluation of Biometric Template Update. In *International Biometric Performance Testing Conference (IBPC 2012)*.
[31] Romain Giot, Christophe Rosenberger, and Bernadette Dorizzi. 2013. A New Protocol to Evaluate the Resistance of Template Update Systems against Zero-Effort Attacks. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. 131–137.
[32] NJ Grabham and NM White. 2008. Use of a novel keypad biometric for enhanced user identity verification. In *Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE*. 12–16.
[33] P. E. Hart. 1968. The condensed nearest neighbor rule. *IEEE Trans. Inform. Theory (Corresp.)* IT-14 (1968), 515–516,.
[34] Mitsutoshi Himaga and Katsuhiro Kou. 2008. Finger Vein Authentication Technology and Financial Applications. In *Advances in Biometrics*, NaliniK. Ratha and Venu Govindaraju (Eds.). Springer London, 89–105.
[35] Syed Zulkarnain Syed Idrus, Estelle Cherrier, Christophe Rosenberger, and Patrick Bours. 2014. Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords. *Computers & Security* 45 (2014), 147–155.
[36] Apple Inc. 2017. Face ID Security. (2017). https://www.apple.com/business/site/docs/FaceID_Security_Guide.pdf
[37] A.K. Jain, A. Ross, and S. Prabhakar. 2004. An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on* 14, 1 (2004), 4–20.
[38] Anil K Jain, Sarat C Dass, and Karthik Nandakumar. 2004. Soft biometric traits for personal recognition systems. In *Biometric Authentication*. Springer, 731–738.
[39] Anil K. Jain, Karthik Nandakumar, and Arun Ross. 2016. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters* 79 (2016), 80 – 105.
[40] Xudong Jiang and Wee Ser. 2002. Online fingerprint template improvement. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24, 8 (2002), 1121–1126.
[41] Pilsung Kang, Seong-seob Hwang, and Sungzoon Cho. 2007. Continual retraining of keystroke dynamics based authenticator. In *Advances in Biometrics*. Springer, 1203–1211.
[42] HB Kekre and VA Bharadi. 2009. Adaptive feature set updating algorithm for multimodal biometrics. In *Proceedings of the International Conference on Advances in Computing, Communication and Control*. 277–282.
[43] Kevin Killourhy and Roy Maxion. 2010. Why did my detector do that?! Predicting keystroke-dynamics error rates. In *Recent Advances in Intrusion Detection*, Somesh Jha, Robin Sommer, and Christian Kreibich (Eds.). Vol. 6307. Springer Berlin / Heidelberg, 256–276.
[44] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. 2011. Activity Recognition Using Cell Phone Accelerometers. *SIGKDD Explor. Newsl.* 12, 2 (2011), 74–82.
[45] Jeffrey W. Lockhart, Gary M. Weiss, Jack C. Xue, Shaun T. Gallagher, Andrew B. Grosner, and Tony T. Pulickal. 2011. Design Considerations for the WISDM Smart Phone-based Sensor Mining Architecture. In *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*. 25–33.
[46] Alessandra Lumini and Loris Nanni. 2006. A clustering method for automatic biometric template selection. *Pattern Recognition* 39, 3 (2006), 495–497.

[47] Sergio Tenreiro Magalhães, Kenneth Revett, and Henrique M. D. Santos. 2005. Password Secured Sites - Stepping Forward with Keystroke Dynamics. In *Proceedings of the International Conference on Next Generation Web Services Practices*. 293–298.

[48] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. 2002. FVC2002: Second Fingerprint Verification Competition. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, Vol. 3. 811–814.

[49] G.L. Marcialis, L. Didaci, A. Pisano, E. Granger, and F. Roli. 2012. Why template self-update should work in biometric authentication systems?. In *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*. 1086–1091.

[50] Aleix Martinez and Robert Benavente. 1998. *The AR Face Database*. CVC Technical Report 24. Centre de Visió per Computador, Universitat Autònoma de Barcelona.

[51] Majid Masso and Iosif I. Vaisman. 2010. Accurate and efficient gp120 V3 loop structure based models for the determination of HIV-1 co-receptor usage. *BMC Bioinformatics* 11, 1 (2010), 1–11.

[52] Abir Mhenni, Estelle Cherrier, Christophe Rosenberger, and Najoua Essoukri Ben Amara. 2019. Analysis of Doddington zoo classification for user dependent template update: Application to keystroke dynamics recognition. *Future Generation Computer Systems* 97 (2019), 210 – 218.

[53] Abir Mhenni, Estelle Cherrier, Christophe Rosenberger, and Najoua Essoukri Ben Amara. 2019. Double serial adaptation mechanism for keystroke dynamics authentication based on a single password. *Computers & Security* 83 (2019), 151–166.

[54] Abir Mhenni, Estelle Cherrier, Christophe Rosenberger, and Najoua Essoukri Ben Amara. 2018. Adaptive biometric strategy using doddington zoo classification of user's keystroke dynamics. In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*. 488–493.

[55] Abir Mhenni, Estelle Cherrier, Christophe Rosenberger, and Najoua Essoukri Ben Amara. 2018. Towards a Secured Authentication Based on an Online Double Serial Adaptive Mechanism of Users' Keystroke Dynamics. In *International Conference on Digital Society and eGovernments (ICDS)*. 73–80.

[56] Abir Mhenni, Estelle Cherrier, Christophe Rosenberger, and Najoua Essoukri Ben Amara. 2018. User Dependent Template Update for Keystroke Dynamics Recognition. In *2018 International Conference on Cyberworlds (CW)*. 324–330.

[57] Abir Mhenni, Christophe Rosenberger, Estelle Cherrier, and Najoua Essoukri Ben Amara. 2016. Keystroke Template Update with Adapted Thresholds. In *International Conference on Advanced Technologies for Signal and Image Processing*.

[58] Denis Migdal and Christophe Rosenberger. 2018. Analysis of Keystroke Dynamics For the Generation of Synthetic Datasets. In *2018 International Conference on Cyberworlds (CW)*. 339–344.

[59] Jugurta Montalvão, Eduardo O. Freire, Murilo A. Bezerra Jr., and Rodolfo Garcia. 2015. Contributions to Empirical Analysis of Keystroke Dynamics in Passwords. *Pattern Recognition Letters* 52, C (2015), 80–86.

[60] Ricardo García Noval and Francisco Perales López. 2008. Adaptative templates in biometric authentication. In *The 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Vol. 2008. 14.

[61] C Pagano, E Granger, R Sabourin, P Tuveri, GL Marcialis, and F Roli. 2015. Context-Sensitive Self-Updating for Adaptive Face Recognition. In *Adaptive Biometric Systems*. Springer, 9–34.

[62] Paulo Henrique Pisani. 2017. *Biometrics in a data stream context*. Ph.D. Dissertation. Universidade de São Paulo (USP) - Instituto de Ciências Matemáticas e de Computação (ICMC).

[63] Paulo Henrique Pisani, Romain Giot, André C. P. L. F. de Carvalho, and Ana Carolina Lorena. 2016. Enhanced template update: Application to keystroke dynamics. *Computers & Security* 60 (2016), 134–153.

[64] Paulo Henrique Pisani, Ana Carolina Lorena, and André C. P. L. F. de Carvalh. 2018. Adaptive Biometric Systems using Ensembles. *IEEE Intelligent Systems* 33, 2 (2018), 19–28.

[65] Paulo Henrique Pisani, Ana Carolina Lorena, and André C. P. L. F. de Carvalho. 2014. Adaptive Algorithms in Accelerometer Biometrics. In *2014 Brazilian Conference on Intelligent Systems (BRACIS)*. 336–341.

[66] Paulo Henrique Pisani, Ana Carolina Lorena, and André C. P. L. F. de Carvalho. 2015. Adaptive Approches for Keystroke Dynamics. In *Neural Networks (IJCNN), The 2015 International Joint Conference on*. 1–8.

[67] Paulo Henrique Pisani, Ana Carolina Lorena, and André C. P. L. F. de Carvalho. 2015. Adaptive positive selection for keystroke dynamics. *Journal of Intelligent & Robotic Systems* 80, 1 (2015), 277–293.

[68] Paulo Henrique Pisani, Ana Carolina Lorena, and André C. P. L. F. de Carvalho. 2015. Ensemble of Adaptive Algorithms for Keystroke Dynamics. In *2015 Brazilian Conference on Intelligent Systems (BRACIS)*. 310–315.

[69] Paulo Henrique Pisani, Ana Carolina Lorena, and André C. P. L. F. de Carvalho. 2017. Adaptive Algorithms applied to Accelerometer Biometrics in a Data Stream Context. *Intelligent Data Analysis* 21, 2 (2017), 353–370.

[70] Paulo Henrique Pisani, Norman Poh, André C. P. L. F. de Carvalho, and Ana Carolina Lorena. 2017. Score normalization applied to adaptive biometric systems. *Computers & Security* 70 (2017), 565 – 580.

[71] Norman Poh, Josef Kittler, and Thirimachos Bourlai. 2010. Quality-based score normalization with device qualitative information for multimodal biometric fusion. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and*

Humans 40, 3 (2010), 539–554.

[72] Norman Poh, Josef Kittler, Chi-Ho Chan, and Medha Pandit. 2015. Algorithm to estimate biometric performance change over time. *IET Biometrics* 4, 4 (2015), 236–245.

[73] Norman Poh, Josef Kittler, Sebastien Marcel, Driss Matrouf, and Jean-Francois Bonastre. 2010. Model and score adaptation for biometric systems: Coping with device interoperability and changing acquisition conditions. In *Pattern Recognition (ICPR), 2010 20th International Conference on*. 1229–1232.

[74] Norman Poh, Joseph Kittler, and Ajita Rattani. 2014. Handling session mismatch by fusion-based co-training: An empirical study using face and speech multimodal biometrics. In *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*. 81–86.

[75] Norman Poh, Joseph Kittler, and Ajita Rattani. 2015. Handling Session Mismatch by Semi-supervised-Based Co-training Scheme. In *Adaptive Biometric Systems*. Springer, 35–49.

[76] Norman Poh, Amin Merati, and Joseph Kittler. 2009. Adaptive client-impostor centric score normalization: A case study in fingerprint verification. In *IEEE 3rd International Conf. on Biometrics: Theory, Applications, and Systems*. 1–7.

[77] Norman Poh, Ajita Rattani, and Fabio Roli. 2012. Critical analysis of adaptive biometric systems. *IET biometrics* 1, 4 (2012), 179–187.

[78] Norman Poh, Rita Wong, Josef Kittler, and Fabio Roli. 2009. Challenges and research directions for adaptive biometric recognition systems. In *Advances in Biometrics*. Springer, 753–764.

[79] Precise Biometrics. 2014. Understanding Biometric Performance Evaluation. (2014). http://precisebiometrics.com/wp-content/uploads/2014/11/White-Paper-Understanding-Biometric-Performance-Evaluation.pdf

[80] Ajita Rattani. 2010. Adaptive biometric system based on template update procedures. *Dept. of Elect. and Comp. Eng., University of Cagliari, PhD Thesis* (2010).

[81] Ajita Rattani. 2015. *Introduction to Adaptive Biometric Systems*. Springer International Publishing, 1–8.

[82] A. Rattani and R. Derakhshani. 2017. Online co-training in mobile ocular biometric recognition. In *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*. 1–5.

[83] A. Rattani, R. Derakhshani, S. K. Saripalle, and V. Gottemukkula. 2016. ICIP 2016 competition on mobile ocular biometric recognition. In *2016 IEEE International Conference on Image Processing (ICIP)*. 320–324.

[84] Ajita Rattani, Biagio Freni, Gian Luca Marcialis, and Fabio Roli. 2009. Template Update Methods in Adaptive Biometric Systems: A Critical Review. In *Advances in Biometrics*, Massimo Tistarelli and Mark S. Nixon (Eds.). 847–856.

[85] Ajita Rattani, Gian Luca Marcialis, Eric Granger, and Fabio Roli. 2012. A dual-staged classification-selection approach for automated update of biometric templates. In *Pattern Recognition (ICPR), 2012 21st International Conference on*. 2972–2975.

[86] Ajita Rattani, Gian Luca Marcialis, and Fabio Roli. 2008. Biometric template update using the graph mincut algorithm: A case study in face verification. In *Biometrics Symposium, 2008. BSYM'08*. 23–28.

[87] Ajita Rattani, Gian Luca Marcialis, and Fabio Roli. 2008. Capturing large intra-class variations of biometric data by template co-updating. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*. 1–6.

[88] Ajita Rattani, Gian Luca Marcialis, and Fabio Roli. 2009. *An Experimental Analysis of the Relationship between Biometric Template Update and the Doddington's Zoo: A Case Study in Face Verification*. Springer Berlin Heidelberg, 434–442.

[89] Ajita Rattani, Gian Luca Marcialis, and Fabio Roli. 2011. Self adaptive systems: An experimental analysis of the performance over time. In *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011 IEEE Workshop on*. 36–43.

[90] Ajita Rattani, Gian Luca Marcialis, and Fabio Roli. 2013. Biometric system adaptation by self-update and graph-based techniques. *Journal of Visual Languages & Computing* 24, 1 (2013), 1–9.

[91] Ajita Rattani, Gian Luca Marcialis, and Fabio Roli. 2013. A multi-modal dataset, protocol and tools for adaptive biometric systems: a benchmarking study. *IJBM* 5, 3/4 (2013), 266–287.

[92] Ajita Rattani, Norman Poh, and Arun Ross. 2013. A Bayesian approach for modeling sensor influence on quality, liveness and match score values in fingerprint verification. In *2013 IEEE International Workshop on Information Forensics and Security (WIFS)*. 37–42.

[93] A. Rattani and A. Ross. 2014. Automatic adaptation of fingerprint liveness detector to new spoof materials. In *IEEE International Joint Conference on Biometrics*. 1–8.

[94] Douglas A Reynolds and Richard C Rose. 1995. Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE transactions on speech and audio processing* 3, 1 (1995), 72–83.

[95] GL Ritter, HB Woodruff, SR Lowry, and TL Isenhour. 1975. An algorithm for a selective nearest neighbor decision rule. *IEEE Transactions on Information Theory* 21, 6 (1975), 665–669.

[96] Fabio Roli, Luca Didaci, and GianLuca Marcialis. 2008. Adaptive Biometric Systems That Can Improve with Use. In *Advances in Biometrics*, NaliniK. Ratha and Venu Govindaraju (Eds.). Springer London, 447–471.

[97] Fabio Roli, Luca Didaci, and Gian Luca Marcialis. 2007. Template co-update in multimodal biometric systems. In *Advances in Biometrics*. Springer, 1194–1202.

[98] Fabio Roli and Gian Luca Marcialis. 2006. Semi-supervised PCA-based face recognition using self-training. In *Structural, Syntactic, and Statistical Pattern Recognition*. Springer, 560–568.

[99] Arun A Ross, Karthik Nandakumar, and Anil K Jain. 2006. *Handbook of multibiometrics*. Vol. 6. Springer Science & Business Media.

[100] Choonwoo Ryu, Hakil Kim, and Anil K Jain. 2006. Template adaptation based fingerprint verification. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, Vol. 4. 582–585.

[101] Tobias Scheidat, Andrey Makrushin, and Claus Vielhauer. 2007. Automatic template update strategies for biometrics. *Otto-von-Guericke University of Magdeburg, Magdeburg, Germany* (2007).

[102] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. 2001. Estimating the Support of a High-Dimensional Distribution. *Neural Computation* 13, 7 (2001), 1443–1471.

[103] F. Schroff, D. Kalenichenko, and J. Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 815–823.

[104] Ashley Willis Scott Goldfine, Rodney Bosch. 2015. Security Sales and Integration. *EH Publishing* (2015).

[105] Mark M Seeger and Patrick Bours. 2011. How to comprehensively describe a biometric update mechanisms for keystroke dynamics. In *Security and Communication Networks (IWSCN), 2011 Third International Workshop on*. 59–65.

[106] Abdul Serwadda, Kiran Balagani, Zibo Wang, Patrick Koch, Sathya Govindarajan, Raviteja Pokala, Adam Goodkind, David-Guy Brizan, Andrew Rosenberg, and Vir V Phoha. 2013. Scan-based evaluation of continuous keystroke authentication systems. *IT Professional* 15, 4 (2013), 20–23.

[107] Anzar S.M., Amala K., Remya Rajendran, Ashwin Mohan, Ajeesh P.S., Mohammed Sabeeh K., and Febin Aziz. 2016. Efficient online and offline template update mechanisms for speaker recognition. *Computers & Electrical Engineering* 50 (2016), 10 – 25.

[108] Rahul Sukthankar and Robert Stockton. 2001. Argus: the digital doorman. *IEEE Intelligent Systems* 16, 2 (2001), 14–19.

[109] Kalaivani Sundararajan and Damon L. Woodard. 2018. Deep Learning for Biometrics: A Survey. *ACM Comput. Surv.* 51, 3 (May 2018), 65:1–65:34.

[110] Matthew E. Taylor and Peter Stone. 2009. Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research* 10 (2009), 1633–1685.

[111] Umut Uludag, Arun Ross, and Anil Jain. 2004. Biometric template selection and update: a case study in fingerprints. *Pattern Recognition* 37, 7 (2004), 1533–1542.

[112] Kumari Vandana. 2007. *Enhancing weak biometric authentication by adaptation and improved user-discrimination*. Master's thesis. International Institute of Information Technology Hyderabad, INDIA.

[113] Zibo Wang, Abdul Serwadda, Kiran S Balagani, and Vir V Phoha. 2012. Transforming animals in a cyber-behavioral biometric menagerie with frog-boiling attacks. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*. 289–296.

[114] Dennis L Wilson. 1972. Asymptotic properties of nearest neighbor rules using edited data. *Systems, Man and Cybernetics, IEEE Transactions on* 2, 3 (1972), 408–421.

[115] David Yambay, Luca Ghiani, Paolo Denti, Gian Luca Marcialis, Fabio Roli, and S Schuckers. 2012. LivDet 2011—Fingerprint liveness detection competition 2011. In *2012 5th IAPR international conference on biometrics (ICB)*. IEEE, 208–215.

[116] Qiuhong Yu, Yilong Yin, Gongping Yang, Yanbing Ning, and Yanan Li. 2012. Face and Gait Recognition Based on Semi-supervised Learning. In *Pattern Recognition*, Cheng-Lin Liu, Changshui Zhang, and Liang Wang (Eds.). Vol. 321. Springer Berlin Heidelberg, 284–291.

[117] Xuran Zhao, Nicholas Evans, and Jean-Luc Dugelay. 2011. A co-training approach to automatic face recognition. In *Signal Processing Conference, 2011 19th European*. 1979–1983.

[118] Xiaojin Zhu. 2005. *Semi-Supervised Learning Literature Survey*. Technical Report 1530. Computer Sciences, University of Wisconsin-Madison.

[119] Xiaojin Zhu, Zoubin Ghahramani, John Lafferty, and others. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, Vol. 3. 912–919.

[120] Indrė Žliobaitė, Albert Bifet, Jesse Read, Bernhard Pfahringer, and Geoff Holmes. 2015. Evaluation methods and decision theory for classification of streaming data with temporal dependence. *Machine Learning* 98, 3 (2015), 455–482.

[121] H. Çeker and S. Upadhyaya. 2017. Transfer learning in long-text keystroke dynamics. In *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*. 1–6.