UNICA

UNIVERSITÀ
DEGLI STUDI
DI CAGLIARI

Ph.D. DEGREE IN

Electronic and Computer Engineering

Cycle XXXV

TITLE OF THE Ph.D. THESIS

Fusion of fingerprint presentation attacks detection and matching: a real approach from the LivDet perspective

Scientific Disciplinary Sector

ING-INF/05

Ph.D. Student:	Marco Micheletto
Supervisor	Prof. Gian Luca Marcialis

Final exam. Academic Year 2021/2022
Thesis defence: March 2023 Session

Dedicato a mio Babbo

Contents

Abstract

The liveness detection ability is explicitly required for current personal verification systems in many security applications. As a matter of fact, the project of any biometric verification system cannot ignore the vulnerability to spoofing or presentation attacks (PAs), which must be addressed by effective countermeasures from the beginning of the design process. However, despite significant improvements, especially by adopting deep learning approaches to fingerprint Presentation Attack Detectors (PADs), current research did not state much about their effectiveness when embedded in fingerprint verification systems. We believe that the lack of works is explained by the lack of instruments to investigate the problem, that is, modelling the cause-effect relationships when two systems (spoof detection and matching) with non-zero error rates are integrated.

To solve this lack of investigations in the literature, we present in this PhD thesis a novel performance simulation model based on the probabilistic relationships between the Receiver Operating Characteristics (ROC) of the two systems when implemented sequentially. As a matter of fact, this is the most straightforward, flexible, and widespread approach. We carry out simulations on the PAD algorithms' ROCs submitted to the editions of LivDet 2017-2019, the NIST Bozorth3, and the top-level VeriFinger 12.0 matchers. With the help of this simulator, the overall system performance can be predicted before actual implementation, thus simplifying the process of setting the best trade-off among error rates.

In the second part of this thesis, we exploit this model to define a practical evaluation criterion to assess whether operational points of the PAD exist that do not alter the expected or previous performance given by the verification system alone. Experimental simulations coupled with the theoretical expectations confirm that this trade-off allows a complete view of the sequential embedding potentials worthy of being extended to other integration approaches.


Acknowledgements

Marco Micheletto gratefully acknowledges the Sardinian Regional Government for the financial support of her/his PhD scholarship (P.O.R. Sardegna F.S.E. - Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2014-2020 - Axis III Education and training, Thematic goal 10, Investment Priority 10ii), Specific goal 10.5.

I want to express my sincere gratitude to my supervisor, prof. Gian Luca Marcialis, for his guidance and support throughout my doctoral research. His invaluable insights, feedback, and encouragement have been instrumental in shaping my work and helping me achieve my academic goals.

I am also thankful to all my colleagues at PraLab, for creating a stimulating and supportive research environment that has facilitated my academic growth and development.

I would also like to thank the reviewers of my thesis, Prof. Julian Fierrez and Prof. Martin Drahanský. A separate thanks to Martin Drahanský for welcoming me during my great stay in Brno.

Finally, I would like to express my special gratitude to Prof. Anil Jain for his dedicated effort in reviewing my thesis. His insightful feedback and constructive criticism have been invaluable in improving the quality of my research and enhancing its contribution to the field. I am deeply grateful for his time and expertise and the opportunity to have benefited from his vast knowledge of the subject matter.

1. Introduction

The ability of detecting fingerprint presentation attacks [1, 2, 3] is also called fingerprint liveness detection, or fingerprint anti-spoofing. It has been boosted in the last ten years thanks to the availability of data sets of spoof and alive fingerprint images which allow to partially overcome the lack of information about the problem. In particular, the International Fingerprint Liveness Detection competition, known as LivDet, is a biennal appointment for academies and companies to make the point about the potentials of state-of-the-art fingerprint liveness detectors or PADs. Fig. 1 points out the main achievements of LivDet from 2011 to 2021, whose results are described extensively in [4].

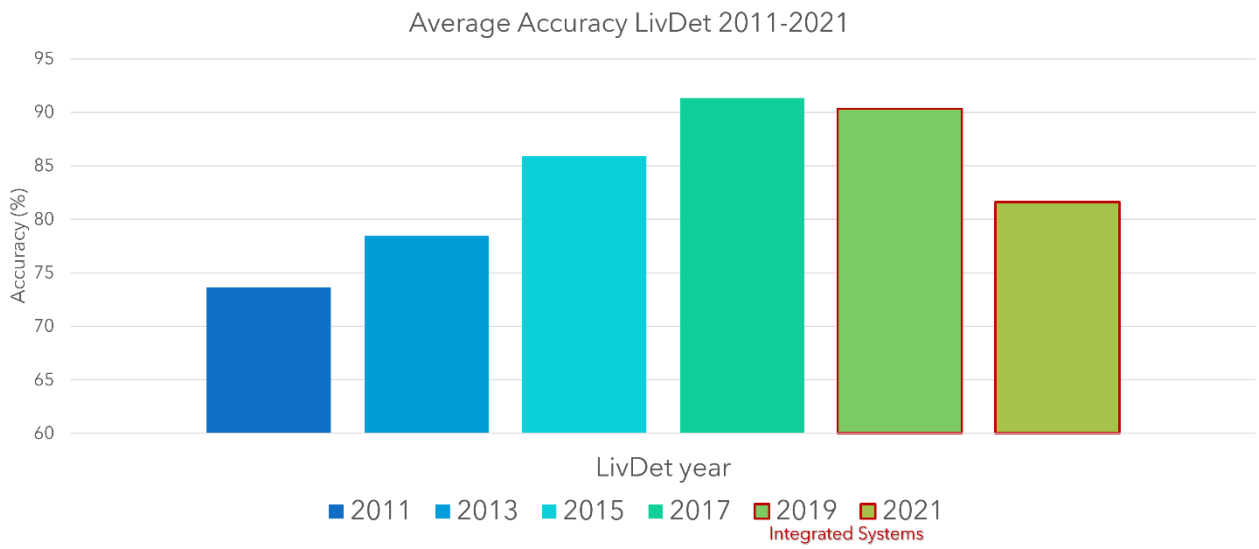


Figure 1.1: Percentage mean accuracy over the datasets and participants to 2011-2021 editions of LivDet [4]

Starting from the LivDet 2019 edition, we began to evaluate not only liveness detection systems, but also their integration with a matching system for personal recognition (Figure 1.2). Apparently in contrast with the trend of the previous competition results (2011-2017), the last two pointed out a sort of "stop" on the increasing average detection rate of PADs, that is particularly marked for the last edition. Although the datasets used in the last competition are more challenging in terms of spoof detection, the observed decrease is not attributable only to this. As a matter of fact, when a PAD is integrated into a fingerprint verification system, several unsolved limitations arise. The most evident one is the apparent gain of attacks detection rate (APCER), at the expense of a loss of genuine acceptance rate (GAR) [5, 6, 7], according to the ISO terminology [8]. This means that a genuine user could be rejected, instead of a match failure, because his/her fingerprint is misclassified as a fake one [9, 10].

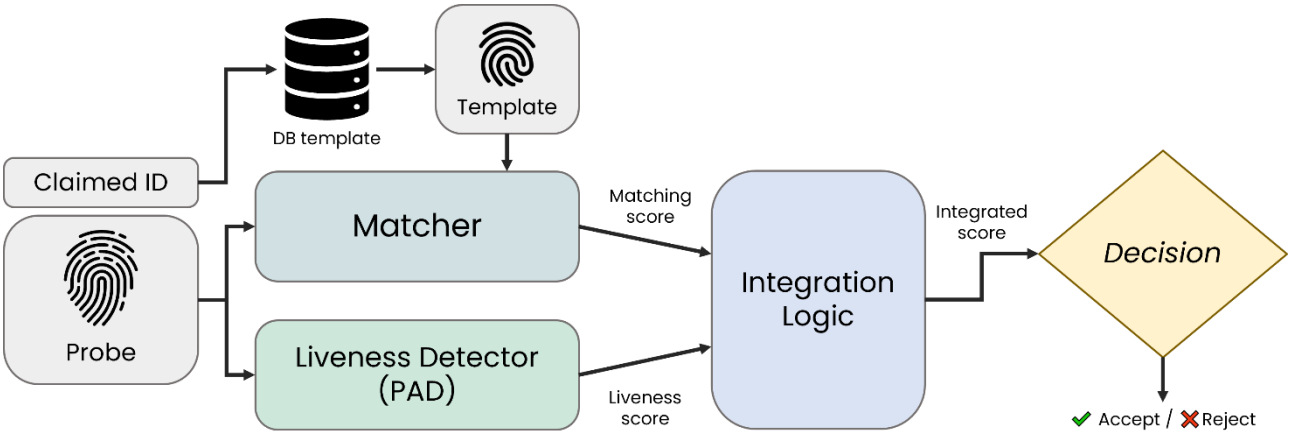


Figure 2.2: Block diagram of a general integrated verification system.

The literature experimentally investigated the integration problem by considering both parallel and sequential combinations of the presentation attack module and the matching module [5, 6, 7, 11, 12, 13]. Recently, the 2nd edition of the Handbook of Biometric Anti-spoofing reported an evaluation methodology for integrated systems, which is also adopted in this work (Chp. 12, [14]).

All of the aforementioned works recognized a performance decrease when PAD and matcher are embedded [5, 9]. The extent of this degradation, however, remains unclear. Can we still rely on the integrated system for security applications? In our opinion, the fact that this error rate amount seriously impacts on the whole system's performance does not make aware about the security breaches that this might lead, once the system has been scaled up for large user populations. This problem appears to be commonly accepted by the community, and considered as intrinsic of a combination of two non-zero error free systems.

One of the core points is the impossibility of doing many things without collecting data, replicating algorithms, and performing experiments. We cannot predict the possible amount of error rates and the conditions of effectiveness of the PAD integration, according to, for example, the adopted sensor, the spoofing materials or the probability of being attacked [14]. Nevertheless, while the sensor characteristic and material adopted are the main variables to evaluate a PAD system, the impact of the spoofing attack probability was never considered at all. For example, high-security and consumer applications (e.g. sensors integrated in smartphones) have different design goals and should be treated diversely: in the second, suppliers may assume that presentation attacks are substantially less probable or irrelevant and accordingly tailor the PAD. However, this is currently not possible. There is no other method to assess current PAD performance when incorporated into verification systems than the ones described above.

Things could change if a tool modelling the presentation attack probability were available. We might determine for which operational points and conditions to implement the given embedding. The tool may evaluate the influence of latex-made PAIs[1] and sensor technology on ROC. Designers would have a practical instrument for assessing high-security and consumer applications.

1.2 Contribution

The first goal of my PhD research was to address the lack of theoretical and experimental explanation of the integration problem, that is, model the cause-effect relationships when two non-zero error-free systems work together. For this purpose, we proposed a novel investigation instrument [15]: a performance simulator of the probabilistic relationships among variables at hand in the case of the sequential fusion of presentation attacks detector and matcher. Sequential fusion is only a possible choice, but it is also the simplest and widespread one.

We called this simulator BIO-WISE and made it publicly available[2]. It takes as input the receiver operating characteristic (ROC) curve of the fingerprint matcher and that of the PAD. The output is the whole acceptance rate in its three basic components: the genuine acceptance rate (GAR), the false non-match rate (FNMR), and the presentation attack acceptance rate (IAPMR)[8]. The simulation can be performed according to two main parameters: the prior probability of being attacked by spoofs , and the specific operational point chosen for the fingerprint PAD.

We did not implement or replicate any PAD algorithms or matching system. This is also what a designer would prefer to do: take the vendors' individual ROCs and explore the performance achievable according to some expected scenarios. Due to the very high detection rate achieved, it is reasonable to consider the PAD algorithms of 2017-2019 LivDet competitions as good representatives of the state of the art. In particular, they represent the performance achievable when detecting presentation attacks by using convolutional neural networks, which are largely acknowledged as the PAD systems of "novel generation". We computed their ROC curves, and investigated their limits and potentials from a theoretical viewpoint during integration with the well-known NIST Bozorth3, that is, the main benchmarking matcher publicly available[3] and the Verifinger 12, namely, the top-level matcher nowadays off-the-shelf[4]. After verifying the reliability of our model in terms of difference between expected and real values, we carried out an extensive set of

[1] Presentation Attack Instrument or, simply, spoof or fake fingerprint.
[2] https://livdet.pythonanywhere.com/
[3] https://www.nist.gov/services-resources/software/nist-biometric-image-software-nbis
[4] https://www.neurotechnology.com/verifinger.html

simulations, explicitly designed to assess the impact on the GAR depending on the materials used for the attack and the probability of being attacked. Accordingly, we derived the main guidelines to follow for deciding whether a particular PAD can be embedded in the matching process.

However, using the current version of BIO-WISE is difficult to appreciate for which PAD's operational points the overall GAR degradation can be still acceptable, with the advantage of handling presentation attacks.

Therefore, as a second contribution of this thesis, we present a significant improved version of BIO-WISE, by introducing a formal definition of "*trade-off*" a term that is used when referring to "a balancing of factors all of which are not attainable at the same".

Through this tool we could quantify the loss of performance in terms of GAR and the related gain in terms of IAPMR and FMR, with particular reference to the best operational points of the liveness module by setting that of the matcher (e.g. the Equal Error Rate (EER)). Using this tool with multiple PADs and matchers can also assess "where we are" based on current technology.

Chapter 2 of this thesis gives an overview of biometric technologies, highlighting the limits and potentials of such approaches. Chapter 3 introduces fingerprint biometry and related works for fingerprint recognition and liveness detection. In Chapter 4, we present Bio-WISE, a novel simulator able to predict the integrated system performance from the individual ROC curves of PAD and verification systems. Chapter 5 present an additional tool, the "trade-off", that significantly improves the capabilities of Bio-WISE. Chapter 5 concludes this manuscript by providing a summary of the major contributions.

2. Biometric technologies

The term "biometric recognition" refers to the process of measuring human biological and behavioral features for the purpose of automatically recognizing or authenticating an individual [16]. It is the science that establishes a person's identity based on **physical** characteristics, such as the fingerprint or palm of the hand, the geometry of the hand or fingers, facial features, iris or retina, or **behavioral** , that is the characteristics influenced by the personality of the individual, such as the style of typing on the keyboard, the gait or the voice. In general, any human characteristics which meet the requirements of universality, uniqueness, permanence and collectability [17] can be used as a biometric identifier.

Biometric recognition can be seen as a particular application of *pattern recognition*. Pattern recognition is a branch of artificial intelligence that deals with locating and interpreting physical (or behavioral) phenomena through the introduction of classification criteria. A *pattern* is, therefore, an information element (a handwritten document, a movement, a sound, a somatic characteristic, etc.) of any nature that can be acquired through an external sensor.

Therefore, creating a **biometric recognition system** means building a system capable of translating an analogue physical phenomenon into a digital description that the machine can understand: once the user's biometric traits have been acquired, they must be transformed into an abstract representation by extracting the feature set. Then, this set is compared with the templates' features stored in a database. The output of this comparison is the so-called ***match score***, which is a measure of the similarity between the two feature sets. Typically, biometric systems can operate in two modes [18] (Fig. 2.2-2.3):

- **Authentication** or **verification**: the system verifies the user's identity by comparing the acquired biometric trait with the corresponding reference, called template, stored in the database. In this mode, the user declares his identity with a PIN, a smart card, a user name, etc.. It is a 1:1 comparison.
- **Identification**: the system compares the input data with all those stored in the database, to determine the most similar. A series of possible candidates is returned, ordered according to the match score. It is a 1:N comparison.

The **enrollment** phase, in which a user's biometric information is collected and saved, precedes both modes of operation. A number or string is used to link the user's identity to the stored data, called **templates** (Fig. 2.1).

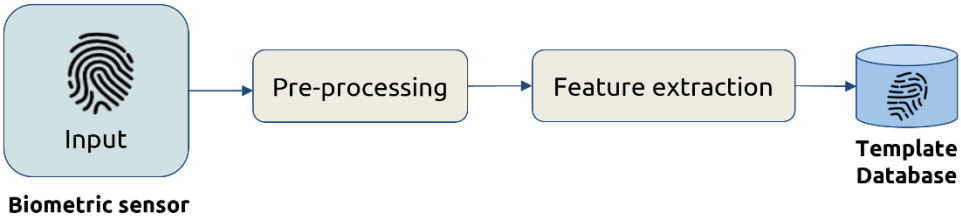


Figure 2.1: Scheme of the enrollment phase of a fingerprint-based biometric system.

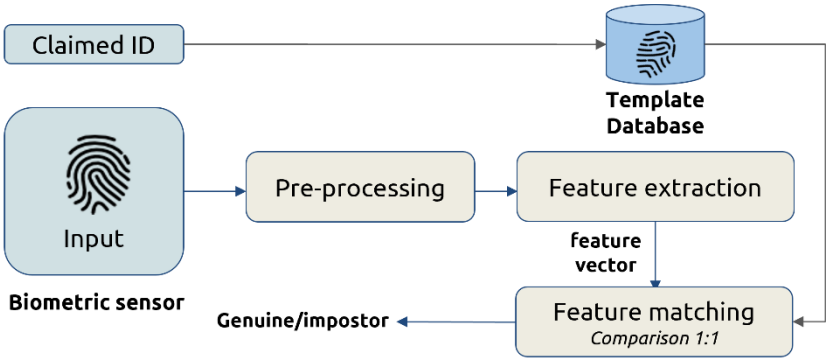


Figure 2.2: Scheme of the authentication mode of a fingerprint-based biometric system.

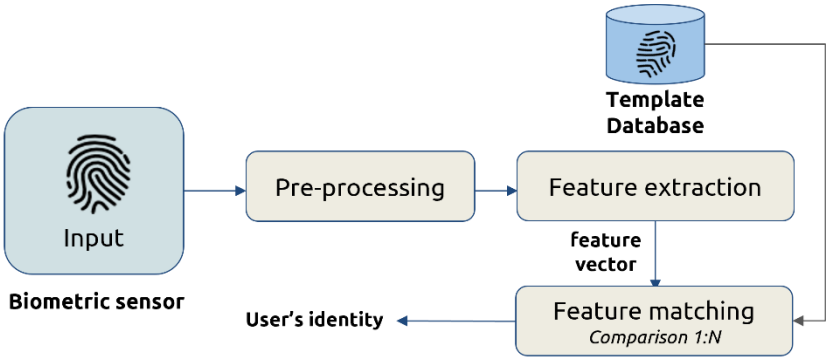


Figure 2.3: Scheme of the identification mode of a fingerprint-based biometric system.

Biometric systems consist of several modules. We generally distinguish:

• **Acquisition module**: responsible for the acquisition and digitization of biometric data. The desired biometry is captured by an electronic device (camera, scanner, fingerprint reader, etc.). It can be

accompanied by a sample quality control module, which discards any templates that could generate unreliable scores in the subsequent stages.

• **Pre-processing or segmentation module**: optional module that improves raw data quality to extract more robust features.

• **Features extractor**: allows the processing of the input image and extracts the distinctive components;

• **Matching module**: outputs the similarity scores (identification) or the score (verification) generated by comparing the input data with those stored in the database. The higher the score, the more similar the input data and the enrolled template.

• **Database**, where biometric templates are stored during the enrollment phase.

2.1 Biometric system limitations

Although biometrics are now commonly used, it is also true that they suffer from some limitations concerning large-scale applications. Depending upon the different scenarios, a biometric system can present certain drawbacks [19]:

- *Performance limitation*: biometrics are systems that base their power on statistics, therefore on probabilistic decisions, which are error-prone. In a verification system, errors are due to many reasons, such as variations in human characteristics (e.g., occlusions [20]), environmental factors (e.g., illuminations [21]) and cross-device matching [22].
- *Architecture limitation*: Between the acquisition module and the comparison module, there are numerous points of vulnerability in which an attack can occur and consequently compromise the security of a biometric system. We can mainly identify eight weak points [20], schematized in Figure 2.4:
    1. **Presentation Attack**: an artificial reproduction of the biometric feature is presented as input to the system. Examples include a fake finger, a copy of a signature, or a face mask.
    2. **Replication of the biometric signal**: the sensor is bypassed, and a recorded signal is replayed to the system, such as an old copy of a fingerprint image or a previously recorded audio signal;

3. **Feature overriding**: the feature extraction module is attacked by a trojan horse in which the attacker preselects the features;
4. **Feature replacement**: the set of extracted features is replaced with a different fraudulent one;
5. **Matcher corruption**: the matcher is corrupt, and the attacker forces the decision;
6. **Template replacement**: one or more templates are modified in such a way that a fraudulent template corresponds to an authorized identity;
7. **Interception of communication**: the template transmitted by the database is intercepted while sent to the matcher and corrupted;
8. **Overriding the final decision**: If the hacker can override the final match decision, then the authentication system has been disabled.

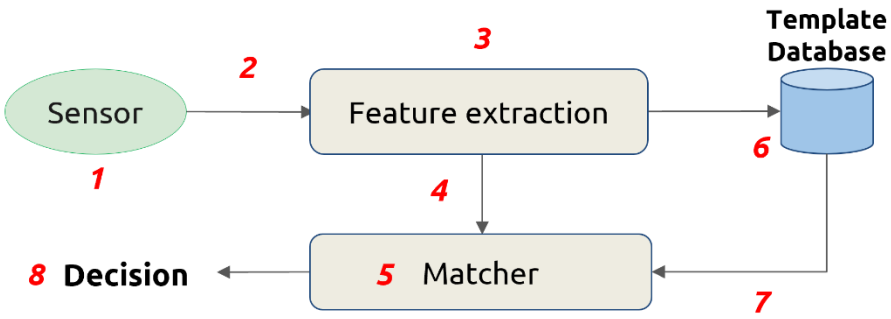


Figure 2.4: Vulnerable points of attacks in a biometric system.

In this PhD thesis, we will focus on the so-called Presentation Attacks (PA) and, in particular, on the impact of the approaches for detecting and preventing such attacks on fingerprint systems, also known as Fingerprint Presentation Attack Detectors (FPADs), when integrated into verification system. This topic will be deepen in the next Section.

2.2 Evaluation of biometric systems

As shown in the previous section, biometric systems have several limitations which may significantly decrease their use in real-life applications. Therefore, the evaluation of biometric systems is carefully considered in the literature. In general, a biometric system can be seen as a **binary classification system** subject to two types of errors:

- False positives (FP): access will be allowed to unauthorized users. It represents a security concern;

- False negatives (FN): access will be denied to authorized users. It is not a security issue.

Typically, the error rates are reported in terms of False Positive Rate (FPR), which corresponds to the ratio between FP and the total number of negative samples and False Negative Rate (FNR), which corresponds to the ratio between FN and the total number of positive samples. Often, the FNR is substituted by its opposite, the True Positive Rate (TPR) intended as the ratio of correctly classified positives.

To compute these error rates, the system needs to be tailored with a decision threshold which will serve as a boundary between the output scores of the genuine accesses and presentation attacks. By changing this threshold, the balance between FPR and FNR will also change: increasing FPR reduces FNR and vice-versa. However, it is often desired that an optimal threshold is chosen according to some criterion. A well-established criterion is, for instance, the Equal Error Rate (EER) [26] defined as:

$$EER = \frac{FPR(s^*) + FNR(s^*)}{2}$$

where s^* is an optimal threshold value for which $FPR = FNR$.

Usually, classification systems are also evaluated via graphical representations of the results. One of the most commonly used method to summarize a system's performance is the Receiver Operating Characteristic (ROC) (Fig. 2.5) . The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The Area Under ROC Curve (AUC) value is often used as a yardstick in order to compare several system: the higher the AUC the better the system.

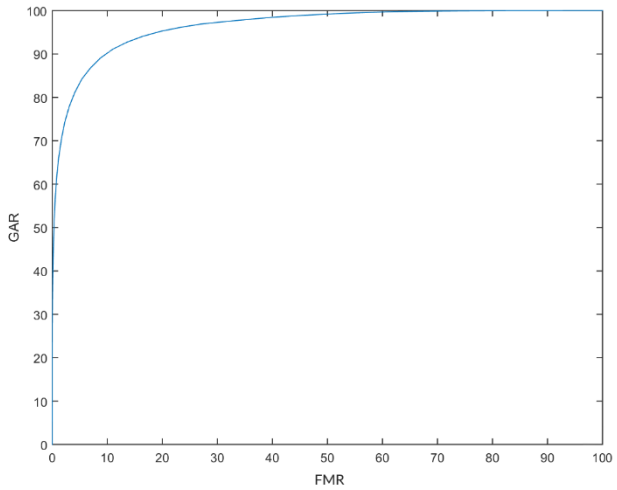


Figure 2.5:Example of Receiver Operating Characteristic (ROC) curve.

2.2.1 Evaluation of verification systems

In the scope of biometric verification systems, positive samples are called **genuine** samples, namely users that are being matched against the same reference identity, while users matched incorrectly against another identity, called **zero-effort** or **impostor** samples, are considered negative samples. Moreover, **False Match Rate** (FMR) and **Genuine Acceptance Rate** (GAR) are the most commonly used terms for FPR and TPR [8, 24]. FMR stands for the ratio of incorrectly accepted zero-effort impostors and GAR for the ratio of correctly accepted genuine users.

2.2.2 Evaluation of presentation attack detection systems

The need to protect against presentation attacks (PAs) led to Presentation Attack Detection (PAD, also known as anti-spoofing). A biometrics system can be designed to automatically detect when a presentation attack is occurring and take appropriate defensive actions against it. Regardless of the technique, biometric mode or degree of independence of external equipment, also PAD systems are most commonly treated as binary classification systems [14]. The two classes that they differentiate are bona-fide[5] (positive samples) and presentation attack samples (negative samples). From this perspective, their evaluation is equivalent to the previously introduced evaluation standards for the binary classification systems. The performance metrics for a PAD has been recently renamed by the ISO/IEC 30107-3 standard [24]: the FPR was retitled to **Attack Presentation Classification Error Rate** (APCER), while the FNR to **Bona Fide Presentation Classification Error Rate** (BPCER). They represent, respectively, the percentage of bona-fide and presentation attacks misclassified by the PAD.

2.2.3 Evaluation of verification systems under presentation attacks

Presentation attack detection systems in biometrics are seldom imagined to operate independently. Their task is to conduct an additional check on the decision of biometric verification systems in order to identify a fraudulent user who owns a replica of a genuine user's biometric feature. As illustrated in Section 2.2.1, verification systems are designed to decide between two categories of verification

[5] Bona-fide are also called real or live samples. Both genuine and zero-effort impostor samples are bona-fide samples. While zero-effort impostors are negative samples in a verification system, they are considered positive samples in a standalone PAD system (since they are not PAs).

attempts: bona-fide genuine users (positives) and the so-called bona-fide zero-effort impostors (negatives) [8]. Presentation attacks represent a new type of samples that can be presented at the input of this system. Considering that both presentation attacks and zero-effort impostors need to be rejected, it is still possible to regard the problem as a binary classification task where the genuine users are the positives, while the union of presentation attacks and zero-effort impostors are the negatives. Nevertheless, tuning of different properties of the verification system to make it more robust to presentation attacks may require a clearly separated class of presentation attacks. Presentation attacks, therefore, should be considered as a third separate category of samples that the verification systems need to handle [25]. This viewpoint, casts biometric verification into a pseudo-ternary classification problem.

Therefore, the usual **False Match Rate** (FMR) and **Genuine Acceptance Rate** (GAR) are coupled with the term **Impostor Attack Presentation Match Rate** (IAPMR), representing the ratio of incorrectly accepted presentation attacks [24]; This will be the terminology adopted in the remainder of this thesis.

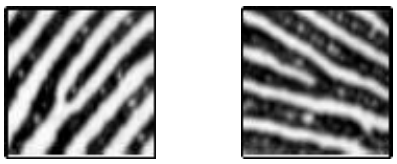
3. Fingerprint Biometric Systems

A fingerprint is an impression left by the friction ridges of a human finger's tip (Fig. 3.1) [27]. This biometric characteristic can reasonably be considered the most mature from both an academic and an industrial point of view. It is not only used worldwide for forensic purposes but also in civilian applications (e.g. smartphone market) thanks to its property of uniqueness and persistence [28]. Fingerprints can be represented by using global information (e.g., finger ridges) or local information (characteristics derived from the ridges).

Figure 3.1: Fingerprint image. The alternating of ridges and valleys on the surface of fingertips generates a unique pattern.

Ridge details are generally described in hierarchical order at three different levels:

- At a global level (Level 1), ridges assume a distinctive shape characterized by high curvatures. These regions, called **singularities** or singular points, can be classified into three types: cycles, deltas and spirals.
- At a local level (Level 2), the details consist of different anomalies referred to as **minutiae** points, representing a discontinuity of the ridge/valley structure. The most important classes of minutiae are ridge endings and bifurcations. A ridge ending generates when a ridge abruptly breaks off, while a bifurcation is the point where a ridge splits into two branches (Figure 3.2).
- At an even finer level (Level 3), if the fingerprint image is acquired at high resolution (1000 dpi), details such as sweat **pores** and incipient ridges can be detected in the fingerprint pattern



(a) (b)

Figure 3.2: Examples of minutiae: ridge ending (a) and ridge bifurcation (b).

The following sections will provide an overview of the modules that compose a typical fingerprint-based biometric system. Subsequently, attention will be focused on the problem of the so-called liveness detection, i.e. the identification of spoof fingerprints. Finally, we will discuss the integration of these two modules, which is the central theme of this thesis.

3.1 Fingerprint recognition

The architecture of a fingerprint-based verification system is typically composed of five modules [27]:

- **A device for image acquisition**: the user's fingerprint is acquired by a fingerprint scanner to produce a raw digital representation;
- **A preprocessing or segmentation module** in which the input image is enhanced to extract distinctive features more easily;
- **A feature extraction algorithm**, which extracts a set of discriminatory characteristics from the acquired and processed image. It is the core of the process;
- **A matcher**, in which the actual recognition process occurs. The input fingerprint is compared against one or more existing templates, returning a score: the higher it is, the more similar the two images are;
- **A database**, where the templates of approved users of the biometric system, also called clients, are usually stored.

In the last decades, several fingerprint recognition algorithms have been developed [29]. They can be coarsely classified into three families:

- *Correlation-based approaches*: two fingerprints are overlapped, and the correlation between pixels for various alignments is computed (for example, by varying the position and orientation) [30,31].
- *Minutiae-based approaches*: this is the most popular and used technique, inspired by the oldest manual procedure for comparing two fingerprints. It consists of the search and subsequent memorization of the minutiae in a feature vector that includes different characteristics such as type, orientation, position, etc. The comparison is essentially based on finding the best alignment between the stored template and the set of input minutiae, which results in the largest number of coupled minutiae [32, 33]. The two matchers employed in this thesis are minutia-based: the NIST Bozorth3 and Verifinger 12.
- *Non-Minutiae feature based approaches*: minutiae extraction is harrowing for extreme low-quality fingerprint images. In these cases, fingerprint pattern properties such as ridges' local direction and frequency, shape, texture information, etc., can be exploited to build a descriptor characterized by compactness, accuracy and robustness [34, 35, 36]

3.2 Liveness detection

Since 1998, when the weakness of personal identification systems to fingerprint replicas was first established [23], the scientific community has studied and proposed novel approaches to defend such systems using so-called Presentation Attack Detectors (PADs). Similarly, attackers' abilities have continued to improve over time, and the study of attack approaches is the core of research on Presentation Attack detection. In this respect, the most common methods for fabricating a spoof fingerprint can be classified into two categories: **consensual** and **non-consensual**.

In the consensual approach, the user's finger is pressed on an impressionable material's surface or wrapped by it, as in the example in Figure 3.3. The negative of the fingerprint is thus fixed on the cast that can now be used as a mold and filled with casting substances, such as latex, liquid ecoflex, or glue. The solidified material is then separated from the mold and represents a replica of the genuine fingerprint, allowing a presentation attack against a fingerprint recognition system [37]. Effective spoofs have also been generated through 2D and 3D sophisticated printing techniques [38]. Cooperative approaches are considered as the worst-case scenario since the user's collaboration permits the creation of a high-quality fake.

Figure 3.3: Creating a mold with silicone rubber (RTV).

Non-consensual methods, on the other hand, depend on recovering indirectly the fingerprint. Latent fingerprints from a smooth or non-porous surface can be exploited after being visualized, since they are not directly visible in most cases.

Various methods to perform this visualisation step are known from forensics. One of the methods is application of very fine-grained powders on the latent fingerprint or take a photo of a smartphone screen [41]. The visualised latent fingerprint is then digitised and converted to a black and white mask that is used in further steps. In this case, the digital version of the mark is used to create the mold. A

laser printer or photolithographic techniques can be used to print it on a transparent sheet, where the spoof materials are then casted (Figure 3.4). Another standard procedure consists of dripping the material after engraving the negative of the fingerprint on a printed circuit [42].

Although such techniques need a high level of precision, good manual dexterity, image processing skills and especially time [50], they involve greater risks for the user as they are more realistic in terms of attacks on a biometric system.

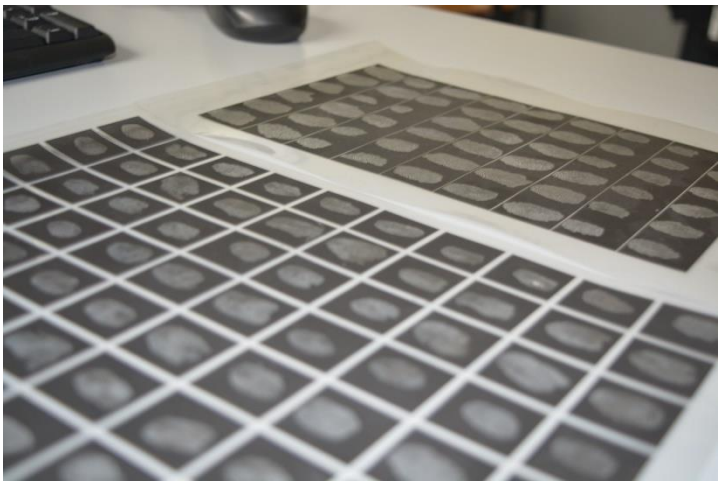


Figure 3.4: Fingerprints obtained with non-consensual method printed on a transparent sheet. They can be used as a cast for creating a spoof.

To defend against these types of attacks, various software solutions for liveness detection have been implemented over the years [39].

Usually, these methods are divided into two categories:

- *Hardware-based techniques*: specific hardware devices are added to the sensor to detect physical traits that ensure liveliness (e.g. heartbeat, blood pressure, odour [43, 44, 45]). These solutions usually have a higher fake detection rate but at the expense of a higher cost and invasiveness, as happens, for example, with blood pressure measurement.
- *Software-based techniques:* the sample's liveness is verified after a standard sensor has acquired it, by augmenting the software architecture with a dedicated attack detection algorithm. Distinctive features are extracted from the image and not from the finger itself. In general, software-based methods are less expensive and, intuitively, non-invasive since their function is transparent to the user.

The real challenge in this field is to make traditional readers reliable and robust against presentation attacks through software-based methods. Over the years, many publications focused on showing performance improvement and the number of datasets for this aim enormously increased. As with other biometrics, presentation attack detection techniques have also evolved from the analysis of ridges and valleys to local hand-crafted methods based on morphology, color and texture analysis, such as BSIF and LBP, and more modern deep-learning techniques [46, 47, 48].

Among the several initiatives, the International Fingerprint Liveness Detection Competition (LivDet), organized by the Biometric Unit of the PRALab[6] of the University of Cagliari, is a well-acknowledge biennial meeting aimed at making the point on the limits and perspectives of PADs. The competition occurred and grew in seven editions between 2009 and 2021, offering novel falsification techniques, spoof materials and types of scanners.

In LivDet (2009-2017) and the majority of works in the literature, the problem of fingerprint liveness detection is treated and evaluated as an independent pattern classification problem, namely, it is not yet viewed as a part of a fingerprint verification system. As LivDet 2019 and 2021 editions showed (Figure 1.1) and other previous works noticed, fusing a PAD with an existing personal verification system is not trivial at all: in all cases, a drop in performance has been noticed, due to the increase in the rejected genuine users, that is, the system's GAR.

This brings some questions and doubts about the integration process: what are the advantages and drawbacks if integration is done with the SOA PADs? Is it possible to evaluate quantitatively the conditions under which these PADs can be integrated without degrading the overall performance? Ideally, a good trade-off between the ability to reject zero-effort and presentation attacks (low FMR and IAPMR) and the need to accept genuine users (high GAR) is necessary. This is not easy to obtain, and experiments are obviously necessary to set such a trade-off, if any.

The following Section will analyze the main achievement in the literature regarding integrated fingerprint systems.

[6] Pattern Recognition and Applications Lab, https://pralab.diee.unica.it/

3.3 State of the art of integrated system: a brief review

Although the research on fingerprints PAD is still hot, a few of works has dealt with the problem of its integration with a verification system. Table 3.1 lists such contributions. The most relevant points that we can summarize from these references are:

- Proposed integration methods rely on sequential fusion [5] or attempt to model liveness and match scores based on probabilistic correlations between measurements and events in the form of a Bayesian network [6, 7, 12, 13]. While these publications are valuable, they do not provide a comprehensive analysis of the integrated system's relationships between the FMR, GAR, and IAPMR. The amount of false rejections is simply reported as a value, different from one system to another, and it is even impossible to predict it a priori.
- In all cases, IAPMR >> FMR for the same acceptance threshold. Consequently, the False Non-Match Rate strongly increases when the individual system is under presentation attack [3, 1, 5] because a more stringent threshold is required to reduce IAPMR. Moreover, experiments in [49] showed that GAR could lower up to 40% when the fingerprint sensor is subjected to presentation attacks. Again, no explanation for this is still given.
- The performance of current PADs discourages integration and suggests other approaches, such as hardware-based liveness detection or tricks like enrolling multiple fingers [50], especially by considering large-scale applications. This cannot be neglected despite the significant efforts of academies and companies to exploit the most recent achievements in machine learning, where the use of deep networks led to a PAD accuracy apparently better than that achieved using hand-crafted features [51]. This can also be observed by looking at Fig. 1.1, where the performance of the 2015 edition is mostly related to the use of hand-crafted features, whilst the one of the 2017 edition is mostly related to the use of deep learning approaches and these are used in almost all algorithms submitted to the 2019-2021 edition [4].

Table 3.1: List of contributions on the integration of PAD and verification systems.

Reference	Contribution	Year
Abhyankar et al. [5]	Integration of fingerprint PAD and verification by sequential fusion. Verification performance is experimentally evaluated. Fusion allows a performance better than that of individual verification system under Presentation Attacks. The impact on GAR of integrating PAD and matcher is noticed.	2009
Marasco *et al.* [6]	Bayesian-based fusion of PAD and verification system. It allows a performance better than that of individual verification system. No theoretical explanation is given about that. The decrease of GAR in order to keep low FMR and IAPMR can be also noticed from the reported results.	2012
Rattani *et al.* [7]	Follow-up of [3]. Several Bayesian-based combinations of liveness and match scores are proposed. A general performance improvement with respect to the individual matcher is pointed out.	2012
Marasco *et al.* [11]	Sequential integration of fingerprint PAD on a multimodal system based on face and fingerprint. Experimental evaluation. No mention about GAR, FMR and IAPMR relations is given.	2011
Rattani and Poh [12]	Joint Density Estimation by Fusion of quality, liveness and matching scores. Performance improvement is pointed out under Presentation Attacks. The contribution of liveness score appears strongly relevant, but no theoretical explanation of the phenomenon is given.	2013
Wong *et al.* [13]	Follow-up of [5]. An extended framework is proposed by including quality measurements as done in [4]. Experimental results show the decrease of GAR in order to keep low FMR and IAPMR.	2014
Biggio *et al.* [49]	A statistical meta-model for security evaluation of multibiometric systems against presentation attacks. IAPMR values are always minor than FMR values. Up to 40% GAR loss is acknowledged for integrated system.	2017
Crossmatch WP [50]	It is pointed out that in order have a reduced probability of being successfully attacked, the user should enrol multiple fingers and then using a randomization procedure of required fingers during the verification stage.	2014

Furthermore, it is unclear how and why the cited integration approaches work and in which cases they could not. In other words, the literature lacks theoretical motivations for what is achieved. The consequence is that quantifying or predicting the advantage of such methods, given the individual performance of PAD and verification system, is currently not possible.

Implications of this drawback are apparent if we consider that the design process needs a phase in which the optimum fusion strategy must be chosen based on the individual systems' performances. Otherwise, the designer could also be interested in simply evaluating the real need for a fusion module, starting, for instance, from the current state of the art and the a priori probability of a presentation attack [14]. However, without a clear model for estimating such performance, all possible approaches must be implemented and tested. Only after this step the selection of the best one is possible according to the working context and related constraints. This is further complicated when evaluating more than one PAD and verification system is necessary.

This work is aimed to start filling this gap by proposing a framework able to present the overall system performance under integration without the implication of the practical difficulties above. The outcome of our research is a simulator of possible scenarios, called Bio-WISE[7], oriented to the sequential fusion of PAD and verification systems. We choose such integration because the link between performance and probabilistic relationships can be modelled easily under appropriate working hypotheses. Moreover, the superiority of more complex approaches to this one has not yet been shown, neither theoretically nor experimentally. Through Bio-WISE, great help can be given to the design phase and to understand to which extent the use of a PAD allows better performance. In particular, the designer can understand whether sequential fusion works depending on the kind of attacks (materials), and the probability of being attacked. Moreover, she/he can state "where we are" according to the current technology on which matchers and PADs are based.

[7] https://livdet.pythonanywhere.com/

4. Bio-WISE: Biometric recognition with integrated PAD: simulation environment

As previously mentioned, the purpose of this study is to present a simulator capable of returning the performance expectation of the sequential fusion of PAD and fingerprint matching. We named this simulator Bio-WISE and made it available to the public. As outlined in Figure 4.1, it accepts as input the ROC curves of the fingerprint matcher and the PAD and returns the integrated system's metrics, namely, the genuine acceptance rate (GAR), the false match rate (FMR) and the impostor attack presentation match rate (IAPMR).In addition, two main parameters are added to perform the simulation: the prior probability of being attacked by spoofs called "w" (see also in Ref. [14]) and the specific operating point chosen for the fingerprint PAD, set by a specific percentage of BPCER or APCER. By adjusting these two criteria, our simulator can represent different case-study to determine when it is most convenient to switch the PAD module on or off.

In the following sections, we will introduce the necessary terminology to illustrate how this simulator is implemented, starting from the appropriate modelling of the probabilistic relationships among the ROC of the two individual systems.

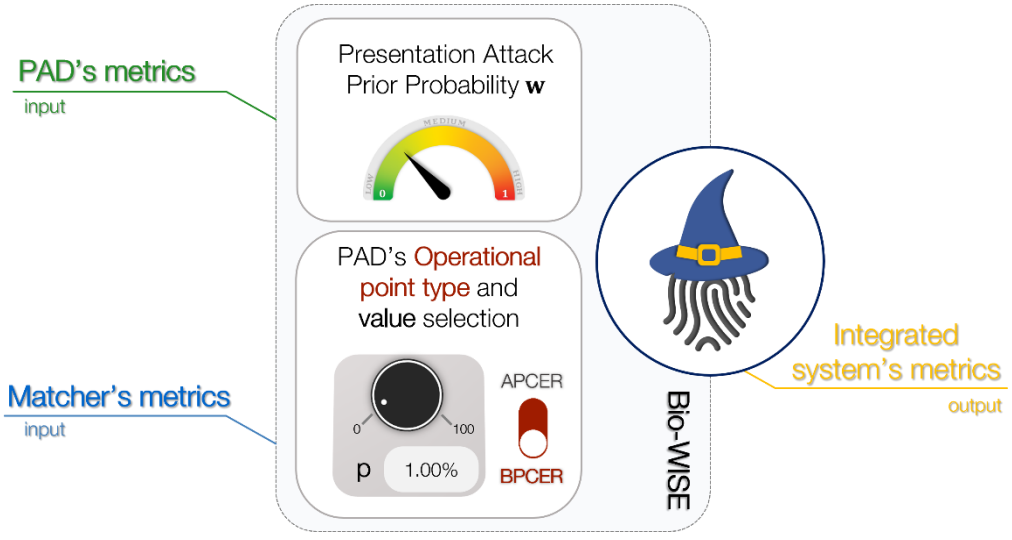


Figure 4.1: Scheme of Bio-WISE, the proposed simulator. It takes the individual metrics of the matcher and the PAD, and return the "performance expectation" of their sequential fusion.

4.1 Problem Modeling

In this Section, we introduce some terms used in the rest of the manuscript. First of all, let G represent the boolean event “the input user is authorized” and $\bar{G}$ the opposite event “the input user is an impostor”. Obviously, $G \cup \bar{G} = \Omega$ where Ω indicates the *Certain Event*. Second, let L represent the boolean event "the input image is alive/authentic", namely, the input picture is from the user's living fingertip. Accordingly, $\bar{L}$ denotes that the input image is from a spoof fingerprint. Even in this case, $L \cup \bar{L} = \Omega$. According to this notation, the following combined occurrences are possible:

- $\{L, G\}$: the input image is alive, and the user is authorized (**genuine user trial**);
- $\{L, \bar{G}\}$: the input image is alive, and the user is unauthorized (**zero-effort attack**);
- $\{\bar{L}, \bar{G}\}$: the input image is spoof and the user is unauthorized (**presentation attack**);
- $\{\bar{L}, G\}$: **impossible event**. An authorized user should never utilize a replica of his fingerprint to access the system.

The relationship between these two variables are summarized by the Venn’s diagram (Fig 4.2). By expressing the probabilities linked to G and L with $P(G)$ and $P(L)$, where $P(G) = 1 - P(\bar{G})$ and $P(L) = 1 - P(\bar{L})$, we can also state:

- the input sample belongs to an unauthorized user, given $L = False$: $P(G|\bar{L}) = 0$;
- the input sample is alive, given $G = True$: $P(L|G) = 1$;
- $P(G, L) = P(G)$.

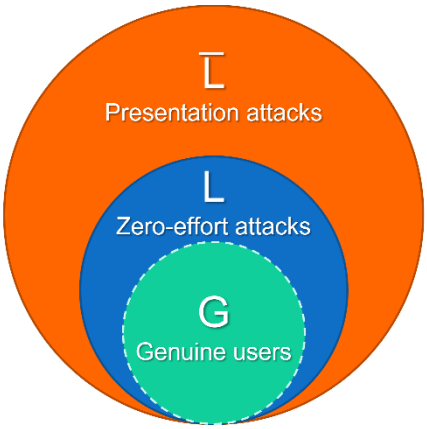


Figure 4.2: Relationship between G and L events.

It is now necessary to link the events L and G, related to the nature of the input sample, to the output of the matching and the liveness detection phases. For this purpose, we introduce two events that appropriately model the acceptance rate of a single matcher and a PAD. As shown in Section 2.2, in a biometric verification system the access is granted to a certain user when the matching score s_M between the input image and the user's claimed identity template(s) is over a given acceptance threshold s_M^*. Consequently, we can define the boolean event M as follow:

$$M = s_M > s_M^* \tag{4.1}$$

Its probability, $P(M)$, represents the acceptance probability of a generic input sample. It can be written in the function of the event L and G exploiting the law of total probability:

$$P(M) = P(M|G,L) \cdot P(G,L) + P(M|\bar{G},L) \cdot P(\bar{G},L) + P(M|\bar{G},\bar{L}) \cdot P(\bar{G},\bar{L}) \tag{4.2}$$

In a similar way we can define the boolean event F, which takes into account the decision of the liveness detector:

$$F = s_F > s_F^* \tag{4.3}$$

F is $true$ if the liveness score s_F, obtained by the analysis of the feature set extracted from the input image, is over a certain liveness threshold s_F^*. Therefore, $P(F)$ is the general probability of classifying a generic pattern as alive:

$$P(F) = P(F|L) \cdot P(L) + P(F|\bar{L}) \cdot P(\bar{L}) \tag{4.4}$$

On the basis of the definitions above we may represent the acceptance rate of each access trial for the individual matcher in terms of the typical error rates used to evaluate its performance:

$$P(M|G,L) = GAR(M) \tag{4.5}$$

$$P(M|\bar{G},L) = FMR(M) \tag{4.6}$$

$$P(M|\bar{G},\bar{L}) = IAPMR(M) \tag{4.7}$$

respectively, the Genuine Acceptance Rate, the False Match Rate and the Impostor Attack Presentation Match Rate.

Likewise, we may depict the bona fide and presentation attack classification error rates of the PAD:

$$P(F|L) = 1 - BPCER(F) \tag{4.8}$$

$$P(F|\bar{L}) = APCER(F) \tag{4.9}$$

4.2 Sequential fusion of liveness detection and matching

The sequential nature of the embedding represented in Fig. 4.3 sets up the final decision to be an AND-like boolean one. This means that an input pattern characterized by a certain state of nature $\{L, G\}$, is finally accepted when both F and M events are $True$. This is valid whether the matcher precedes the PAD and vice versa.

In the following, we model the expression of the acceptance rate by using the terminology previously introduced. We avoid to specify the actual value of truth associated to L and G, since analogous expressions can be obtained for each configurations of these random variables.

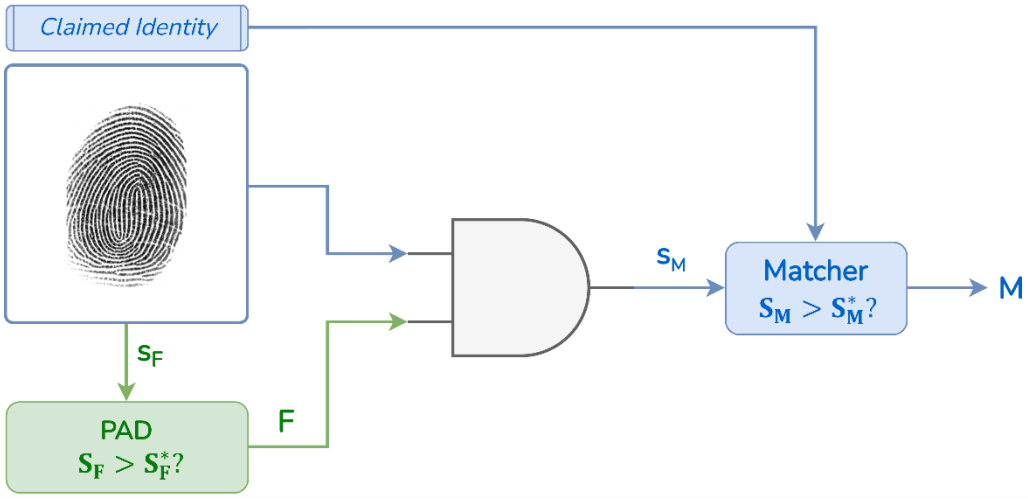


Figure 4.3: Serial combination of presentation attack detector and matcher. The PAD-matcher fusion is a particular case of AND fusion system.

4.2.1 Matching and liveness

Let us first consider an integrated system where the identity of a specific user is verified before the liveness detection. In this case, the probability of acceptance, given a specific configuration of $\{L, G\}$, can be expressed as:

$$P(M, F|L, G) = P(M|L, G) \cdot P(F|L, G, M) \tag{4.10}$$

This formulation can be simplified under the following assumptions:

(1) ***F* is independent of *G*, given *L* and *M*.** No information about the liveness probability is given by the evidence that the sample belongs to an authorized user, since $\boldsymbol{F \neq L}$. Accordingly:

$$P(F|L,G,M) = P(F|L,M)$$

(2) ***F* is independent of *M*, given *L***: again, there is no probabilistic correlation between the liveness score and the fact that the input sample match or not. The decision of the liveness module is independent of the fact that a specific fingerprint matches the corresponding user's template, also according to the definition itself of PAD, which must detect the liveness separately regardless of the user population. Therefore:

$$P(F|L,M) = P(F|L)$$

In summary, if we report the simplified expression of the acceptance rate for all the admissible combinations of L and G, and substitute the system rates to the probabilities (Eqs. 4.5-9), we can observe that the classification error of a "*matcher→PAD"* integrated system is always the simple product of the error rates of the two individual systems:

Probabilities	***Error rates***
$P(M,F\|L,G) = P(F\|L)\cdot P(M\|L,G)$	$\Rightarrow GAR_{Matcher\rightarrow PAD} = GAR(M)\cdot(1-BPCER(F))$
$P(M,F\|L,\bar{G}) = P(F\|L)\cdot P(M\|L,\bar{G})$	$\Rightarrow FMR_{Matcher\rightarrow PAD} = FMR(M)\cdot(1-BPCER(F))$
$P(M,F\|\bar{L},\bar{G}) = P(F\|\bar{L})\cdot P(M\|\bar{L},\bar{G})$	$\Rightarrow IAPMR_{Matcher\rightarrow PAD} = IAPMR(M)\cdot APCER(F)$

(4.11)

Worth noting, according to the extensive literature about the decision-level fusion of multiple classifiers, the sequential fusion of matcher and presentation attacks detector can be treated as a standard AND rule [52]. The AND fusion rule is the logical product of two boolean variable, in this case M and F. However, the literature does not conclude that the classification error of a generic AND-based fusion system is always the simple product of the error rates of the individual systems. Thanks to the formulation above, we are also able to give a prediction of the performance by estimating $BPCER(F)$, $APCER(F)$, $GAR(M)$, $FMR(M)$ and $IAPMR(M)$ independently of each other. Therefore, the proposed model considers the PAD-matcher fusion as a particular case of AND fusion system, where even the error rate evaluation can be treated similarly.

Moreover, we can explain some more interesting relationships between M and F:

- From the point of view of zero-effort attacks, we may see that the GAR/FMR ratio is unaltered, independently of the presentation attacks detector's performance.
- The formulation above implicitly confirms what was experimentally reported in [5]: beyond FMR and $IAPMR$, also the GAR decreases in the integrated system, if we keep the same operating point of the matcher. Such loss is intrinsic to this kind of fusion and it is inversely proportional to the $BPCER$ of the presentation attacks detector.

4.2.2 Liveness and matching

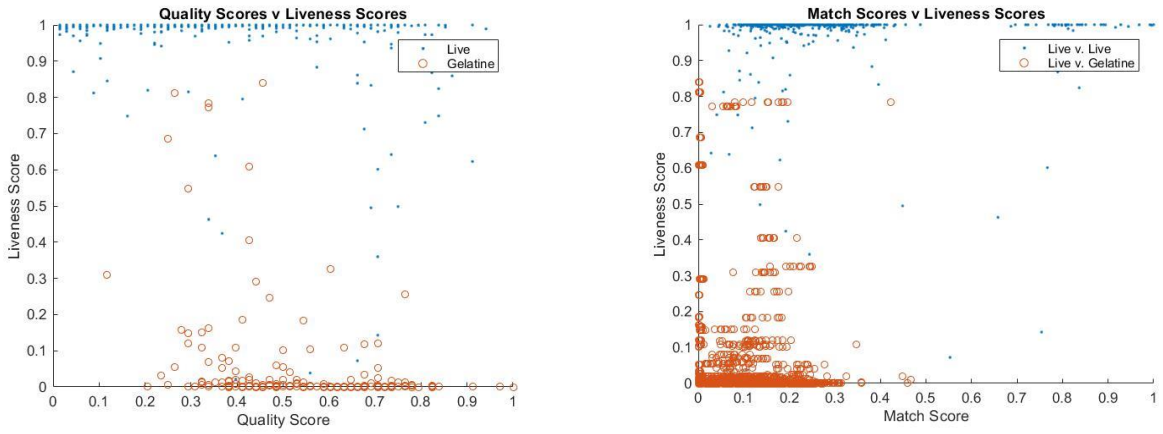
In this architecture, the liveness detection is carried out before the verification of the identity (Fig. 4.3), namely, the input sample is verified only after it is classified as alive. The acceptance rate consequently assumes the following form:

$$P(M, F|L, G) = P(F|L, G) \cdot P(M|L, G, F) \tag{4.12}$$

We can simplify the expression by recalling that ***F* is independent of *G*** (assumption 1 in the previous Section):

$$P(M, F|L, G) = P(F|L) \cdot P(M|L, G, F) \tag{4.13}$$

The last step is determining the degree of dependence between M and F events. We already hypothesized that F is independent of M. However, let us now discuss it more in-depth. The question is: what is the probabilistic dependence of obtaining a match given that the submitted fingerprint image is classified as alive? A recent work [40] investigated the statistical relationships between match scores, quality scores and liveness scores of fingerprint spoofs created from latent fingermarks. Based on the reported results, we can assume that no explicit statistical dependence can be found between the match score and the liveness score. For the sake of example, we report in Figure 4.4 the plots of match and quality scores vs the correspondent liveness scores when a presentation attack is committed by a spoof fabricated with one of the most effective materials, namely, the gelatine. It is evident that there is no significant correlation among such measurements. Therefore we can confirm what we have previously stated: there is no difference in having a spoof or a live fingerprint on the sensor's surface: the probability of a match is unaffected.



(a) (b)

Figure 4.4: Quality/match scores vs liveness scores when the gelatine is used to fabricate PAIs. These plots are quoted from [9], where more details are available.

Consequently:

$$P(M,F|L,G) = P(F|L) \cdot P(M|L,G) \tag{4.14}$$

Hence, the final expressions of GAR, FMR and IAPMR are the same as reported in Eqs. 4.11, regardless of whether the matcher precedes or follows the PAD:

$$\begin{aligned} GAR_{PAD\rightarrow Matcher} &= GAR_{Matcher\rightarrow PAD} = GAR(M) \cdot (1 - BPCER(F)) \\ FMR_{PAD\rightarrow Matcher} &= FMR_{Matcher\rightarrow PAD} = FMR(M) \cdot (1 - BPCER(F)) \\ IAPMR_{PAD\rightarrow Matcher} &= IAPMR_{Matcher\rightarrow PAD} = IAPMR(M) \cdot APCER(F) \end{aligned} \tag{4.15}$$

Hereinafter, we will refer to them as GAR_{Seq}, FMR_{Seq}, and $IAPMR_{Seq}$, respectively.

4.2.3 The final model: *Bio-WISE*

On the basis of the previous Section, and by recalling Eq. 4.2, we obtain the acceptance rate for an integrated system:

$$\begin{aligned} AR_{Seq}(M,F) &= P(M,F) = \\ &= GAR_{Seq}(M,F) \cdot P(G,L) + FMR_{Seq}(M,F) \cdot P(\bar{G},L) + IAPMR_{Seq}(M,F) \\ &\cdot P(\bar{G},\bar{L}) \end{aligned} \tag{4.16}$$

Since $P(G,*)/P(\bar{G}) = P(*|\bar{G})$, we can rewrite Eq. 4.16 as:

$$AR_{Seq}(M,F) = GAR_{Seq}(M,F) \cdot P(G) + GFMR_{Seq}(M,F;w) \cdot P(\bar{G}) \tag{4.17}$$

Where $GFMR_{Seq}$ is the so-called Global FMR, a weighted sum of the two negative errors in function of the term w (see also FAR_{ω} in Ref. [14], p. 471):

$$\begin{aligned} GFMR_{Seq}(M,F) &= FMR_{Seq}(M,F) \cdot P(L|\bar{G}) + IAPMR_{Seq}(M,F) \cdot P(\bar{L}|\bar{G}) = \\ &= FMR_{Seq}(M,F) \cdot (1-w) + IAPMR_{Seq}(M,F) \cdot w \end{aligned} \tag{4.18}$$

Worth noting, the term w of Eq. 4.18 is also reported as a parameter, ω, in Ref. [14], with the following definition: "*ω denotes the relative cost of presentation attacks with respect to zero-effort impostors*". In other words, we have proved that it can also be understood as the prior probability of the system being exposed to presentation attacks since it corresponds to $P(\bar{L}|\bar{G})$. Accordingly, the values of w are between zero and one, and they are representative of the various security scenarios; in particular the higher is w, the higher is the risk of presentation attacks.

Eqs. 4.17-18, coupled with Eq. 4.15, define our simulator. The final Receiver Operating Characteristics (ROCs) can be derived by considering the individual ROCs of the presentation attacks detector and the matcher, with a great help for the designer who is not forced to carry out additional experiments to compute the whole ROC. As a matter of fact, performing experiments with presentation attacks is expensive due to the problem of recovering a user population representative enough and fabricating fake fingerprints, consensually or not. Thanks to Bio-WISE, human effort and cost can be remarkably saved during the design process. Besides this model clearly shows that the GAR decrease is intrinsic to the sequential fusion, other considerations can be done by appropriate simulations, which are carried out in the next Sections. In particular, the GAR decrease is counteracted by the PAD's effectiveness in preventing spoofing attacks and distinguishing them from zero-effort attacks. Furthermore, the prior probability of a spoofing attack w strongly impacts on justifying the addition of a PAD module to the matcher. By acting on w, and the PAD's operational point $BPCER = p\%$ or $APCER = p\%$, the designer may depict several possible scenarios and decide whether the $GFMR_{Seq}$ is better than that of the individual matcher.

4.3 Experiments and analysis

In this Section, we investigate by experiments the Bio-WISE model by some case-studies from the 2017-2019 edition of LivDet.

4.3.1 Datasets

In 2009, to evaluate the main results of the state of the art in the field of fake fingerprint recognition, the Department of Electrical and Electronic Engineering of Cagliari, in collaboration with the Department of Electrical and Computer Engineering of the University of Clarkson, organized the first international liveness detection competition, the Fingerprint Liveness Detection Competition (LivDet) [4]. The competition aimed to compare biometric detection methodologies using a standardized test protocol and a dataset of live and spoof fingerprints created ad hoc by the Biometrics Laboratory of the University of Cagliari for each competition. Approximately every two years, further contests were established to assess the yearly development in this domain. Registration is open to all academic and industrial institutions with a liveness detection solution. These participants are invited to submit their algorithm in an executable which will be analysed and tested on the above dataset to verify its performance. Once the competition is over, the results are published in the related scientific paper.

In the following experimental analysis, we utilized LivDet 2017 and 2019 [4] datasets. They both comprise three datasets of live and spoof fingerprints captured each of them with a different sensor: two optical, GreenBit and DigitalPersona and a thermal swipe, Orcanthus (Figure 4.5). Table 4.1 displays the specific properties of the sensors. The spoofs were generated using several different materials, and always following a consensual procedure. The general distribution of the fingerprint images between both sets is given in Table 4.2 and 4.3.

Table 4.1: Device characteristics for LivDet 2017 and LivDet 2019 datasets

Scanner	**Model**	**Resolution [dpi]**	**Image Size [px]**	**Format**	**Type**
GreenBit	DactyScan84C	500	500x500	BMP	Optical
Orcanthus	Certis2 Image	500	300xn	PNG	Thermal Swipe
DigitalPersona	U.are.U 5160	500	252x324	PNG	Optical

Table 4.2: Composition of the LivDet 2017 dataset.

	Train				Test			
Dataset	Live	Wood Glue	Ecoflex	Body Double	Live	Gelatine	Latex	Liquid Ecoflex
GreenBit	1000	400	400	400	1700	680	680	680
Orcanthus	1000	400	400	400	1700	680	658	680
DigitalPersona	999	400	400	399	1700	679	670	679

Table 4.3: Composition of the LivDet 2019 dataset.

	Train						Test			
Dataset	Live	Wood Glue	Ecoflex	Body Double	Latex	Gelatine	Live	Mix1	Mix2	Liquid Ecoflex
GreenBit	1000	400	400	400	-	-	1700	680	680	680
Orcanthus	1000	400	400	400	-	-	1700	680	658	680
DigitalPersona	1000	250	250	-	250	250	1700	679	670	679

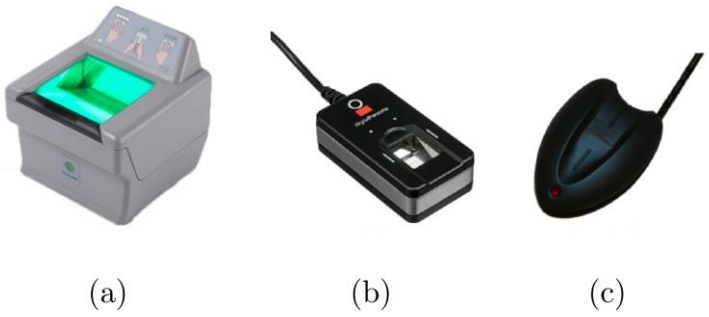


Figure 4.5: Sensors adopted in LivDet 2017 and 2019 editions: GreenBit (a), DigitalPersona (b) and Orcanthus (c).

4.3.2 Experimental protocol

Before employing the model to investigate the actual impact a liveness detector module can bring to a fingerprint verification system, it is necessary to validate the thesis expressed by Eqs. 4.15. For this purpose, we must measure the differences, in terms of performance, between a real sequential integrated system and our simulated scenario, and check if they are comparable. In our evaluation, we considered several PAD systems based both on hand-crafted features and deep learning methods.

In particular, all the algorithms submitted to LivDet 2017 and 2019 competitions were exploited by testing them on relative edition datasets: in the 2017 edition, the solutions adopted are equally distributed between deep learning and hand-crafted algorithms, while in 2019, almost all detectors are based on deep learning approaches. Subsequently, for each dataset, we adopted the following experimental protocol:

1. We computed the liveness score and match score (using the standard Bozorth3 matcher);
2. We computed the individual ROC curves for both PAD and verification system;
3. We estimated the theoretical ROC curve of the sequential integrated system by applying Eqs. 4.15. In our architecture, the liveness module precedes the matcher.
4. We computed the performance of the system separately, without the help of Eqs. 4.15. In other words, we computed the experimental ROC curves according to the standard design approach.

We summarized the results by two possible operational points: a stringent one, where only 1% of presentation attacks can be misclassified, and a relaxed one, where 1% of live samples is incorrectly rejected. Therefore, $APCER(s_F^*) = 0.01$ in the first case, and $BPCER(s_F^*) = 0.01$ in the second one. Of course, all operational points could be investigated. Let us say that the selected ones represent two case-studies quite extreme: in the first case, we may think to a context where it is necessary to assure that very few attacks can be tolerated, due to security constraints; the second case is typical to a service open to a large number of users where it is supposed that a few of attacks are performed over time and it is much more important that no users are "blocked" by the PAD module.

4.3.3 Validation

The following experimental analysis points out the model's reliability in predicting a real sequential system. As mentioned in the previous Section, we investigated over eighty PADs from LivDet 2017 and 2019, computing the absolute difference between significant indexes (FMR, GAR and IAPMR) estimated by Eqs. 4.15 and those obtained through the standard design approach at the selected operational points. For the sake of space, we then extrapolated a set of statistical parameters from the results to show the estimation error, expressed in percentage points, introduced by our model. Since we were not interested in emphasizing the performance differences over the sensors or in assessing the best PAD, we evaluated a global estimate error by computing the average of each acceptance rate over all PADs. We report the mean and standard deviation of such errors in Table 4.4.

Table 4.4: LivDet 2017 and 2019 datasets: Mean and standard deviation of the absolute difference of FMR, GAR, and IAPMR between a standard and an estimated scenario for the two investigated operational points. Reported values are not fractional.

		LivDet 2017	LivDet 2019
$APCER_{01}$	FMR	0.0265 ± 0.023	0.0206 ± 0.023
	GAR	0.9376 ± 0.388	0.2911 ± 0.265
	$IAPMR$	0.0254 ± 0.021	0.0475 ± 0.136
$BPCER_{01}$	FMR	0.0094 ± 0.011	0.006 ± 0.008
	GAR	0.1532 ± 0.046	0.1330 ± 0.092
	$IAPMR$	0.1910 ± 0.150	0.0855 ± 0.133

To better visualize these results, we summarized them with the help of box plots. Figure 4.6 reports the analysis conducted when the PAD is working at the $APCER = 1\%$ ($APCER_{01}$ hereinafter). Firstly, we notice that the two LivDet datasets are similar in terms of error distribution. In both cases, FMR and IAPMR distributions are characterized by a lower variability with respect to the genuine acceptance rate (GAR) one. We can appreciate this graphically since the interquartile range (IQR) of GAR distribution, that is the difference between the third quartile (Q3) and the first one (Q1) (i.e. the total box length), is relatively more extensive than that of SFAR and FAR distributions. This is more marked in the 2019 edition, whose distributions also present a positive skew, namely most of the observations are concentrated on the low end of the scale. Another point to highlight is the presence of outliers. By definition, if a value is outside of the Q3 + 1.5 · IQR or the Q1 − 1.5 · IQR range, that value will be considered an outlier. There are many strategies for dealing with outliers in data, depending on the application and dataset. Fig. 4.6 shows outliers only on SFAR error distribution and analyzing their value, we can state that they have no statistical significance for the purposes of our investigation.

On the other hand, when a very relaxed liveness threshold is set ($BPCER = 1\%$, referred to as $BPCER_{01}$ from now on), the IAMPR estimation error undergoes substantial growth, while GAR and FMR distributions' decrease (Figure 4.7). In this case, we do not identify particular skewness in the distributions; therefore, the probability of getting estimation errors is higher than in the previous instance. Nevertheless, the maximum range is smaller since the largest non-outlier for both editions is relatively lower. This was partly expected since the integrated system performances with a tolerant PAD threshold are pretty similar to those of the verification system alone. Unlike the previous case, outliers also occur in the FMR and GAR error distributions, even though they do not pose a significant danger to the goodness of fit.

In summary, the outcomes of this study indicate that the predicted performances are similar to the experimental data. The most significant error value of this data collection was generated by a 2017 liveness detector in the GAR distribution at the FPR01 operating point, and is around 1.88. In other words, the largest difference between our model and the actual curve is, on average, less than 2%. We may thus infer that our model simulates the sequential combination of a presentation assault detector and a fingerprint verification system with an acceptable level of accuracy.

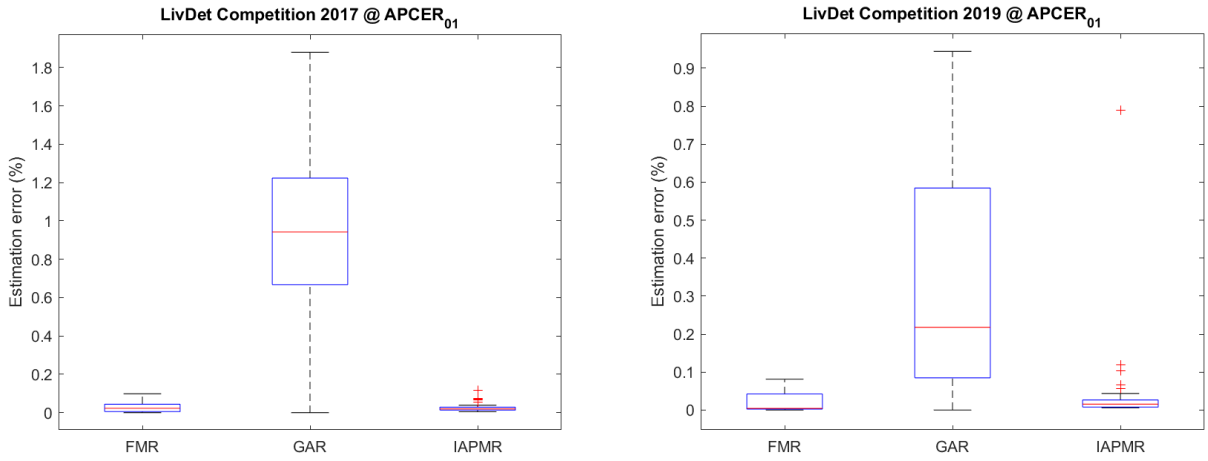


Figure 4.6: Box plots of the absolute difference of FMR, GAR, and IAPMR between a standard and an estimated scenario for the $APCER_{01}$ operational point in LivDet 2017 and 2019.

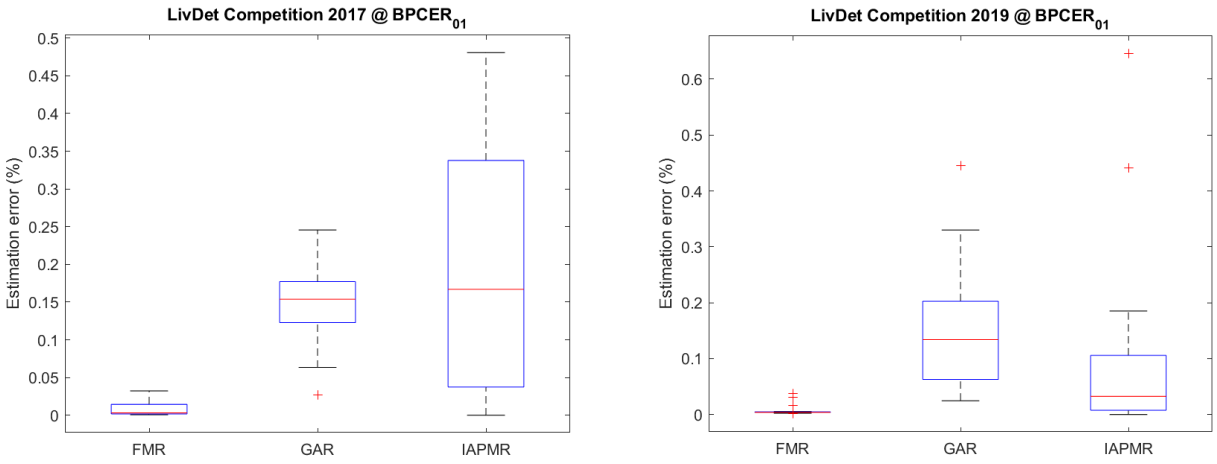


Figure 4.7: Box plots of the absolute difference of FMR, GAR, and IAPMR between a standard and an estimated scenario for the $BPCER_{01}$ operational point in LivDet 2017 and 2019.

4.3.4 Simulations: are we ready for integration?

Once the experimental analysis confirmed the reliability of our simulator in predicting the integrated system performances, we were able to focus on verifying the effect caused by embedding state-of-the-art PAD modules into a state-of-the-art fingerprint verification one.

Section 4.2.3 defines the GFMR index (Eq. 4.18) as a weighted sum of FMR and IAPMR through the term w. Thanks to this term, which represents the PA probability, it is possible to evaluate the system's response to several scenarios, with and without liveness detection embedded. To this purpose, we performed various experiments and plotted the Global ROC curves (GROC, $GFMR$ vs GAR) when $w = [0.00, 0.75]$. Choosing such a large, though unlikely, range allows us not only to test different security scenarios but also to show PADs' behavior in the worst possible case.

In this evaluation, we offer two sets of plots: in the first one, we compare the integrated with the corresponding individual system equipped with the standard matcher **Bozorth3**; in the other, the comparison is carried out by adopting the top-level matcher **Verifinger 12**. PAD side, we analyze all the algorithms submitted to LivDet 2017 and 2019 competitions. Nevertheless, for the sake of space, we propose only the results of the winning algorithm submitted to LivDet 2019 edition, i.e., the algorithm named "*PADUnkFv*" [53], since its behavior is representative of the majority of the presented algorithms. Moreover, to set cross-dataset and cross-material experiments, we tested all the datasets collected in that edition and the previous one.

This analysis led to the identification of two critical scenarios in which the integration significantly degrades the overall performance of personal verification: one involving the acquisition sensor and the other involving the PAI material. In the next Sections, we will analyse in detail such scenarios and propose effective solution to handle them.

4.3.4.1 Sensor dependent analysis

In this section, we compare the performance of the two integrated systems with that of the verification system alone for the LivDet 2019 datasets. The presented ROCs offer a clear overview of the impact of PAD in the two extreme cases investigated in this work. For instance, when the system is adjusted to the $APCER_{01}$ operational point, it is evident that the integration usefulness depends entirely on the probability of attack w. The GAR loss introduced by embedding the PAD is significant (especially for the DigitalPersona sensor, Fig. 4.10a-c) and must be justified by high-risk situations of

presentation attacks, where the probability of PA is very high ($w > 0.5$). Setting such a crucial working point in a relatively low-risk context becomes no longer convenient, given the drop in GAR. As pointed out by Eqs. 4.16, such loss is intrinsic to this kind of fusion and depends on the goodness of the liveness module. Suppose the number of rejected genuine users proves to be outside the design constraints. In that case, the PAD could be avoided or used as a sort of" warning" information during the system's operations. Otherwise, the threshold could be relaxed to obtain a better performance on genuine users but misclassifying more fake fingerprints. This is significant evidence: wherever the meta-designer chooses to operate in a very conservative operational point, he/she can decide to act in advance on the system parameters to improve the performance.

In fact, when genuine acceptance is the first care ($BPCER_{01}$ operational point), the performances are much more balanced. The integrated system does not exhibit a notable $APCER$ value increase by strongly reducing the $IAPMR$ as w rises.

In general, the common aspects highlighted by Figures 4.8-4.10 are mainly two:

1. A system with an embedded PAD is more robust to w variations than a simple matcher. Graphically, we can notice that the curves' dispersion is nearly null, especially when the precautionary threshold is set ($APCER_{01}$). In other words, the performance does not decay if w increases, as is the individual case. This means that the PAD is working correctly and blocking attack attempts from fingerprint forgers.

2. Although Verifinger 12 is a top-level fingerprint verification algorithm, compared with Bozorth3, the absence of substantial performance differences when considering their integration with the PAD suggests that this is the leading "actor" to the global system's effectiveness. We will thoroughly investigate this aspect in Section 4.3.4.3.

A peculiar aspect, instead, is represented by the integrated system working on the Digital Persona sensor. Figure 4.10{a-c} shows the PAD performance, achieved at $APCER_{01}$ operational point.

We immediately notice that the performance drop is considerably higher if compared to the other competition sensors. We can assume that this behavior is due to the acquisition surface. It is significantly reduced compared to the GreenBit sensor, as reported in Table 4.1 (the Orcanthus sensor has a different acquisition technology). This may represent an obstacle in capturing the defects that may appear in the fake fingerprint edge, and that could facilitate PA detection. To support this thesis, we report the average attack presentation classification error rate computed on all LivDet 2019 algorithms (Figure 4.11): the shape of the $APCER$ curve is more relaxed in the case of the DigitalPersona sensor, which means that the percentage of false positives (i.e., the fake fingerprints

classified as alive) is greater under the same threshold. For this reason, the $APCER_{01}$ operational point corresponds to a value of the liveness threshold such as to lead to misclassification of a significant number of live fingerprints.

The role of our simulator is fundamental in this case: If a designer becomes aware of this evidence, he can avoid using this sensor or set another operating point if this scenario does not meet the security constraints of the targeted application. When the PAD is working at $BPCER_{01}$ operational point (Fig. 4.11{b,d}), the integrated system performance improves and gains a fair degree of robustness as the probability of attack increases. As we pointed out in the previous section, the operational points at which both systems operate profoundly impact the final performance since we have the individual error/detection rate product.

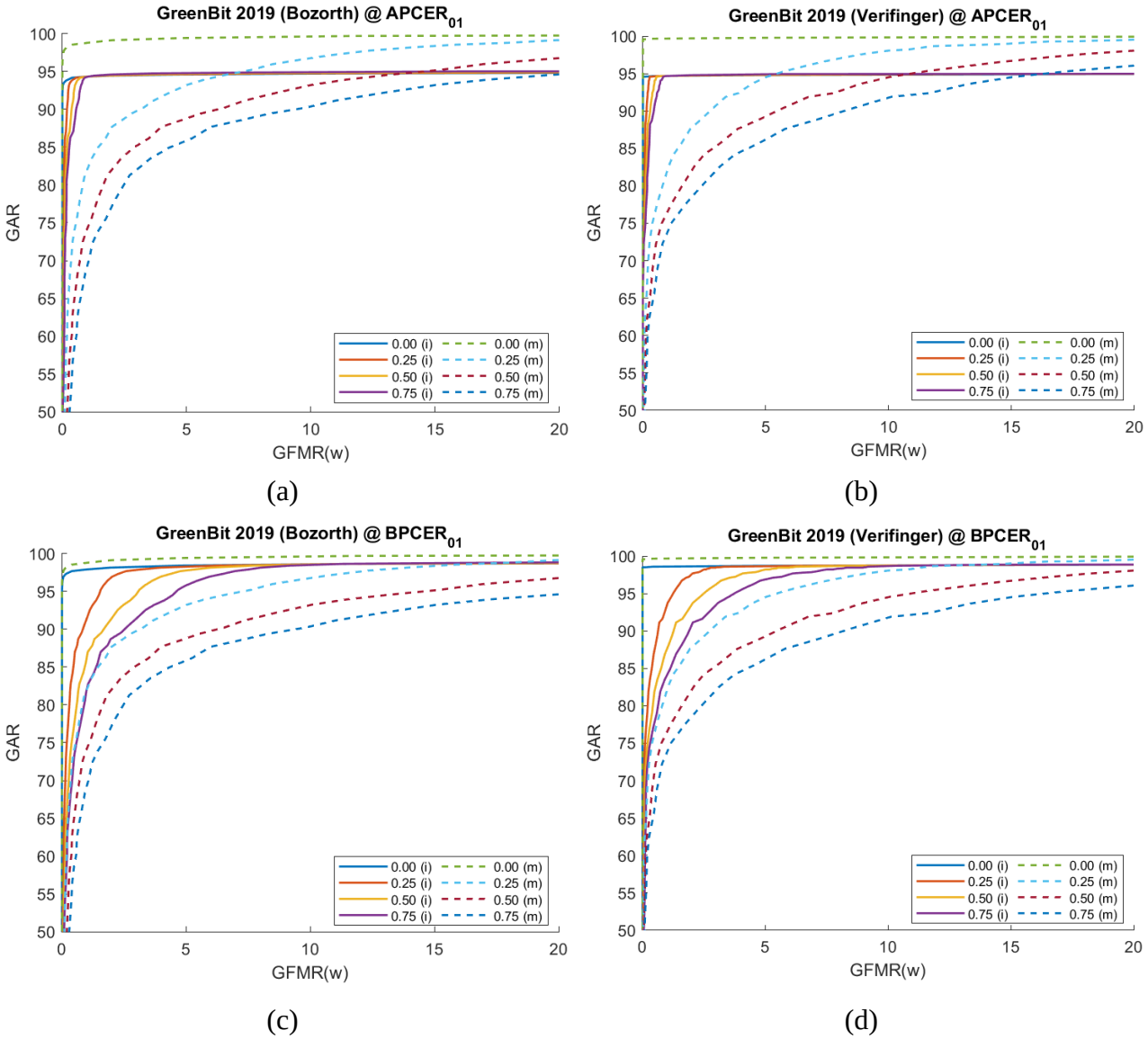


Figure 4.8: GreenBit dataset from LivDet 2019. Comparison between GROCs for an integrated (solid line) and individual (dashed line) matching system equipped with Bozorth3 (a,b) and Verifinger 12 (c,d), varying the presentation attacks probability w. Both operational points are reported for each matcher.

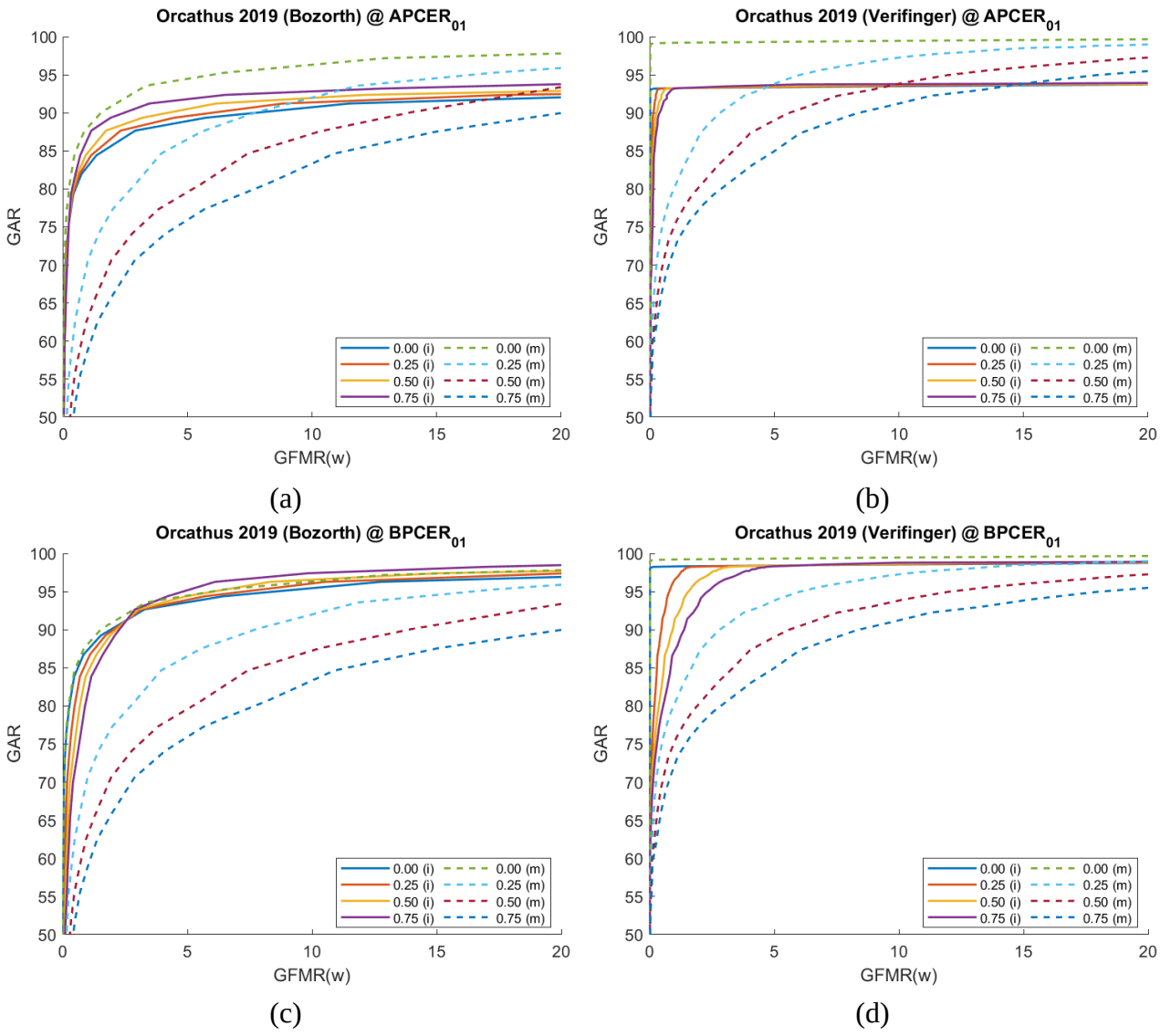


Figure 4.9: Orcanthus dataset from LivDet 2019. Comparison between GROCs for an integrated (solid line) and individual (dashed line) matching system equipped with Bozorth3 (a,b) and Verifinger 12 (c,d), varying the presentation attacks probability w. Both operational points are reported for each matcher.

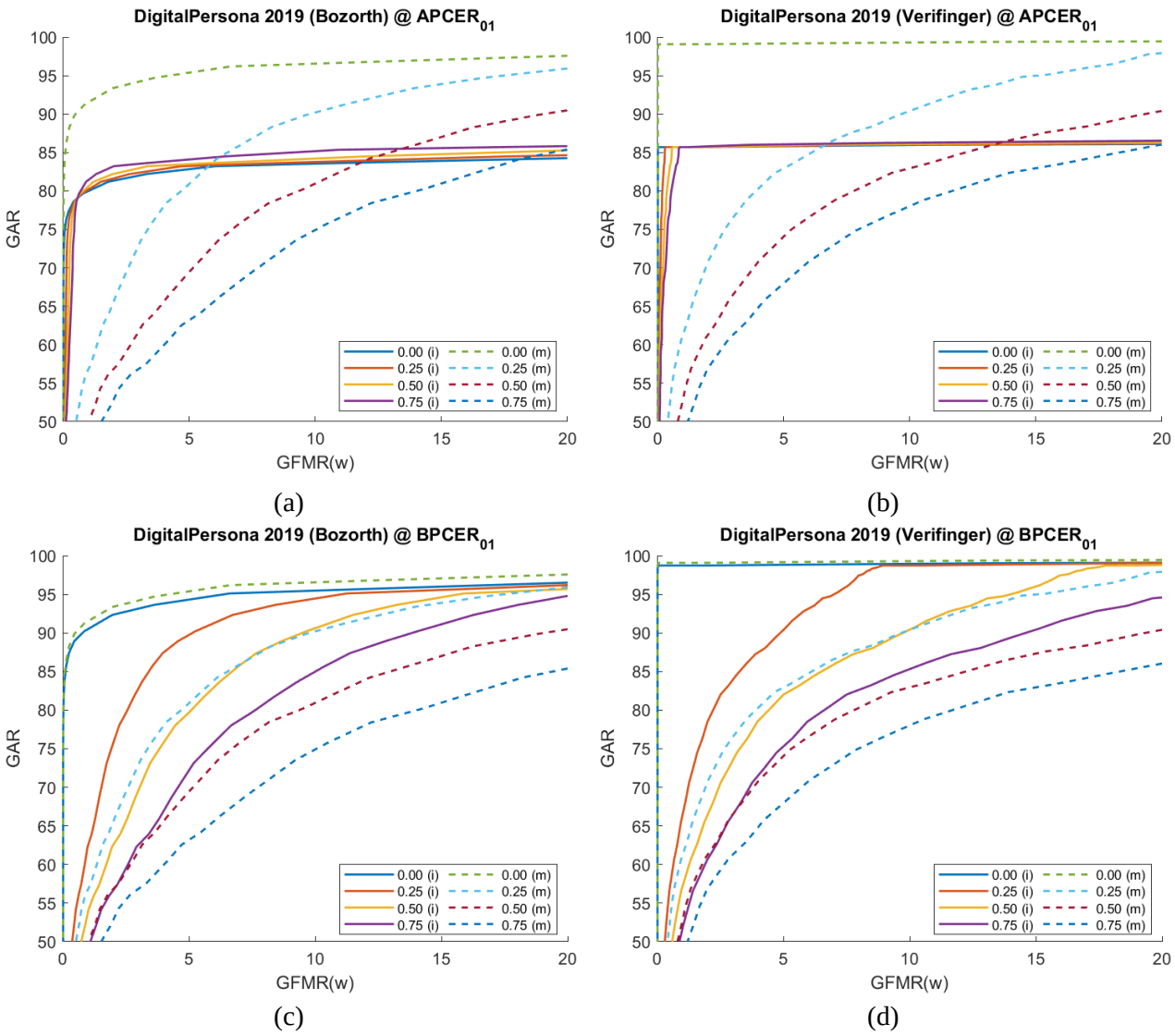


Figure 4.10: DigitalPersona dataset from LivDet 2019. Comparison between GROCs for an integrated (solid line) and individual (dashed line) matching system equipped with Bozorth3 (a,b) and Verifinger 12 (c,d), varying the presentation attacks probability w. Both operational points are reported for each matcher.

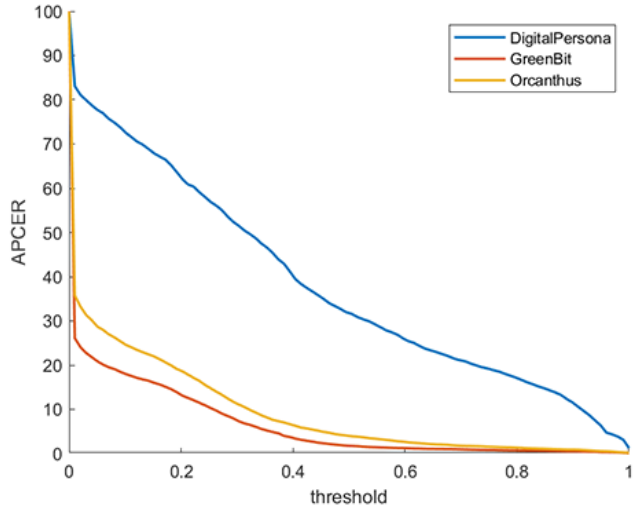


Figure 4.11: Average Attack Presentation Classification Error Rate (APCER) compared in the three sensors of the LivDet 2019 edition. A slow decay trend characterizes the case of the DigitalPersona dataset if compared to the other sensors.

4.3.4.2 Material dependent analysis

In this investigation, we considered the integrated systems obtained with the LivDet 2017 datasets. The three test sets of LivDet2017, as reported in Tables 4.2 and 4.3, differ in both the fingerprints number and composition from those of the 2019 edition. However, the training set is the same regarding the GreenBit and Orcanthus sensors. This lets us figure out which types of materials are best or worst for a specific type of detector. It is commonly acknowledged that the liveness detector reacts differently depending on the spoof material [2, 10]. What kind of effects does this have on the whole system? How do the GAR and GFMR vary? If the PAD behavior is inconsistent with all materials, then an attack with a particular material may completely change the design expectations. It is important to remember that this research aims not to determine which system has the highest performance but to highlight the information that a designer might use during the preliminary stages of a project. Bio-WISE can effectively demonstrate their influence on the system where sensitive materials exist.

Similar to the previous section, we evaluated the LivDet 2019 winner on the LivDet 2017 datasets. When the liveness threshold is set to $APCER = 1\%$ value, we observed a substantial decrease in classification accuracy compared to the corresponding 2019 datasets (Fig. 4.12{a,c} and 4.13a,c}). Since we are examining the same acquisition sensors, in this case, we can only attribute this behavior to the different materials utilized in the two competitions.

Therefore, we conducted further analyses at the classification level for each material. Figure 4.14 illustrates the distinction between the two above cases: LivDet2017's gelatin-made spoofs are erroneously classified as alive fingerprints in a percentage higher than other materials. This means that gelatin can reproduce the natural fingerprint's characteristics so well that it can fool the detector more frequently than any other material in the two datasets.

In addition, the percentage difference between the two indices associated with common materials ("L. Ecoflex" and "Latex") is significantly lower than that between "Mix" and "Gelatine." If a more relaxed threshold is chosen, such as $BPCER = 1\%$, the performance of the integrated system tends to improve, especially as w grows. Therefore, the presentation attacks detector must be studied concerning critical materials before being selected. In summary, we can say that there are still unresolved questions about some materials since the best LivDet2019 detector cannot reach a satisfactory performance on them. This holds for gelatin, one of the most widely accessible and inexpensive materials.

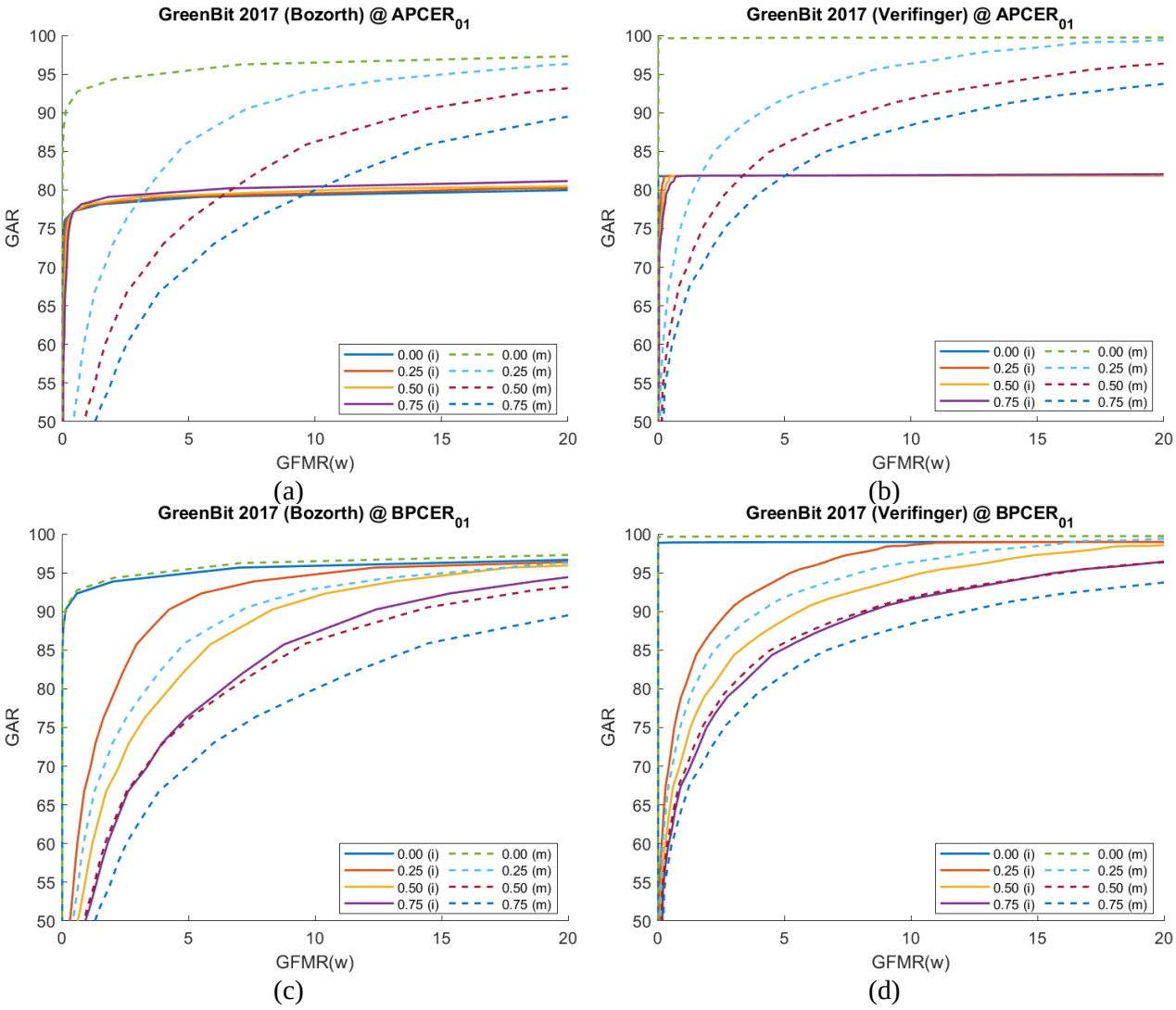


Figure 4.12: GreenBit dataset from LivDet 2017. Comparison between GROCs for an integrated (solid line) and individual (dashed line) matching system equipped with Bozorth3 (a,b) and Verifinger 12 (c,d), varying the presentation attacks probability w. Both operational points are reported for each matcher.

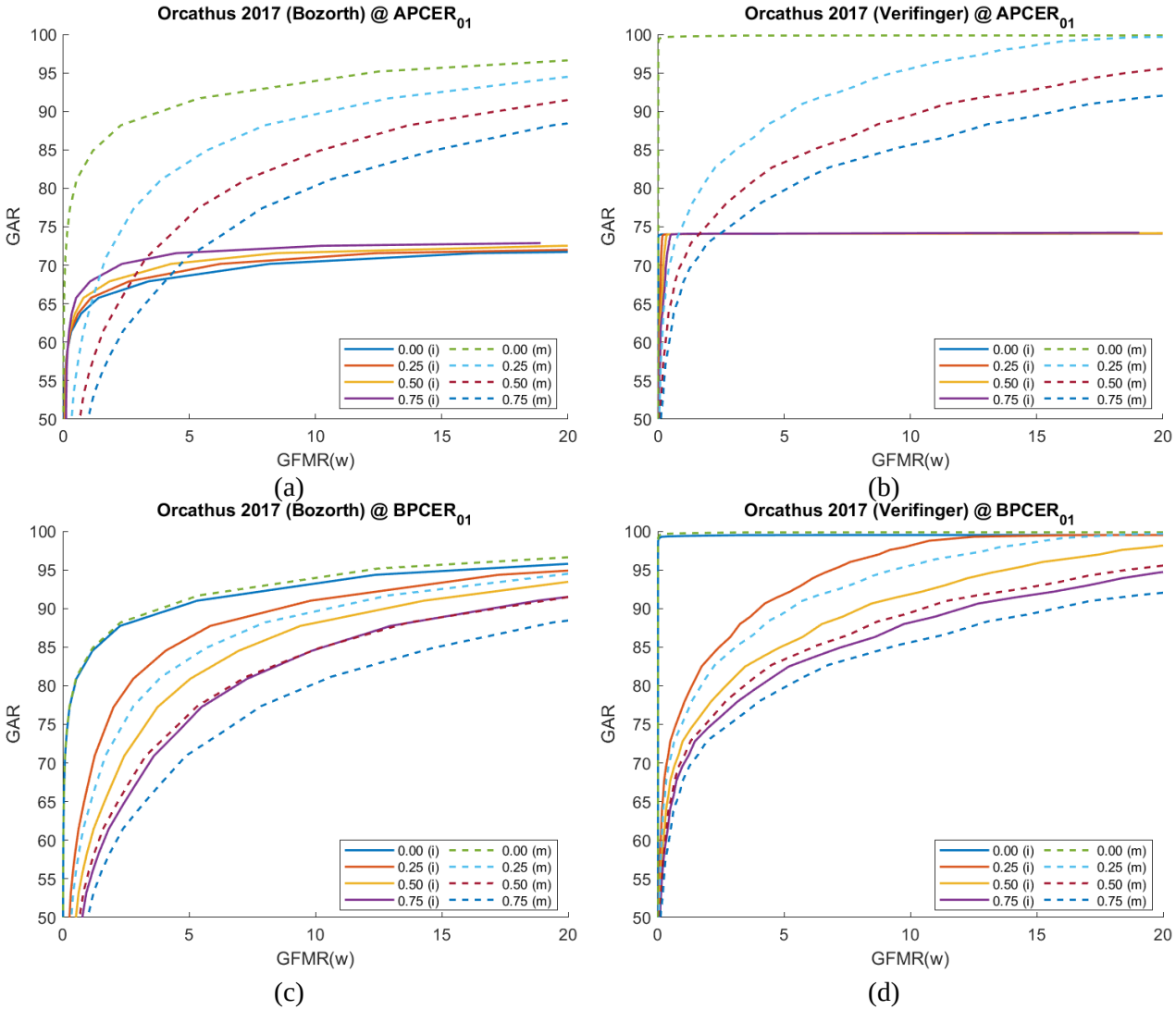


Figure 4.13: Orcanthus dataset from LivDet 2017. Comparison between GROCs for an integrated (solid line) and individual (dashed line) matching system equipped with Bozorth3 (a,b) and Verifinger 12 (c,d), varying the presentation attacks probability w.
Both operational points are reported for each matcher.

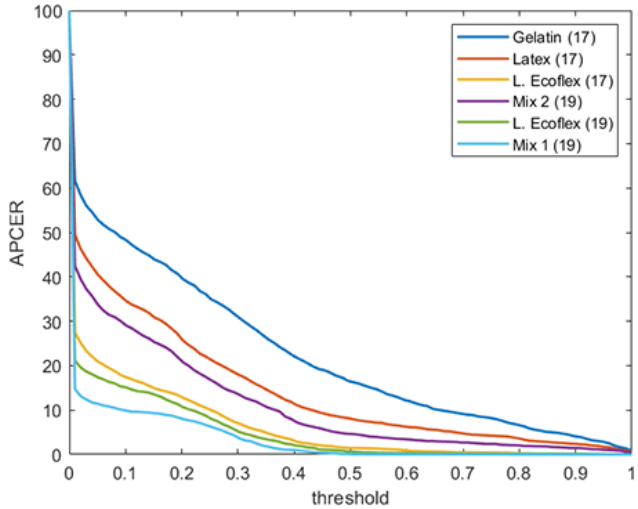


Figure 4.14: Average APCER compared for the GreenBit sensor of LivDet 2017 and 2019.
The "Gelatin" material used in LivDet 2017 originates fake fingerprints harder to classify.

4.3.4.3 PAD dependent analysis

A last investigation is proposed to corroborate the hypothesis about the PAD weight in the integrated system's efficiency. The graphs reported so far suggest that the crucial phase during the design phase lies in the liveness detector's choice.

To corroborate this assumption explicitly, we present experimental results attesting to the behavior of two distinct PADs. In particular, we compared the 2019 LivDet winner with the runner-up, the "*JLW LivDet*" algorithm. In Table 4.5, we highlight the primary differences between these two algorithms by providing their results in terms of $BPCER@1\%APCER$ and $APCER@1\%BPCER$ on all datasets from the 2019 edition of LivDet. We evaluated both algorithms using the most challenging sensor in the competition, the DigitalPersona. However, our conclusions apply to all evaluated datasets. Similar to the previous experiments, we report the results obtained by varying both the PAD operational point (BPCER01 or APCER01) and the matcher (Bozorth3 or Verifinger 12).

Figure 4.15 shows the outcome of the comparison. For each plot, we depict two attack scenarios: one with a moderately low risk ($w = 0.25$) and one with a severe risk ($w = 0.75$). As can be seen, they confirm the thesis expressed up to now. Since the variations between graphs 16a, c and 16b, d are similar, the degradation in overall system performance is due to the different PAD types and is independent of the matcher. This effect is enhanced as the liveness threshold becomes more restrictive.

To be fair, investigated matchers are all minutiae-based. Based on the scientific and technological SOTA, this technology is the most reliable and mature. We cannot say whether different results may be achieved using different matchers based on other features (textural, filters). Examining the impact of the integrated system of deep learning-based matchers rather than minutiae-based is complex and out of the scope of this paper. An additional support of our findings is that PAD systems are based mainly on the training-by-example approach, the accuracy of which depends on several factors, such as the training set representativeness, avoiding overfitting, etc. As a matter of fact, these systems are greatly influenced by the input pattern, which may exhibit significant differences concerning those adopted for training (never-seen-before attacks) and generate very unexpected responses. Therefore, the related performances are generally less robust than matchers, where representativeness is strictly defined in terms of the unicity of the subject’s fingerprint.

Table 4.5: LivDet 2019 dataset. Comparison of BPCER@1%BPCER and APCER@1%BPCER for the two most accurate liveness detectors of the LivDet 2019 competition.

		GreenBit		**DigitalPersona**		**Orcanthus**	
Name	Type	BPCER@1% APCER	APCER@1% BPCER	BPCER@1% APCER	APCER@1% BPCER	BPCER@1% APCER	APCER@1% BPCER
PADUnkFv	Deep learning	5.00%	14.22%	14.03%	40.95%	5.96%	5.88%
JLW_LivDet	Hand-crafted	0.39%	0.33%	26.67%	55.60%	3.23%	5.51%

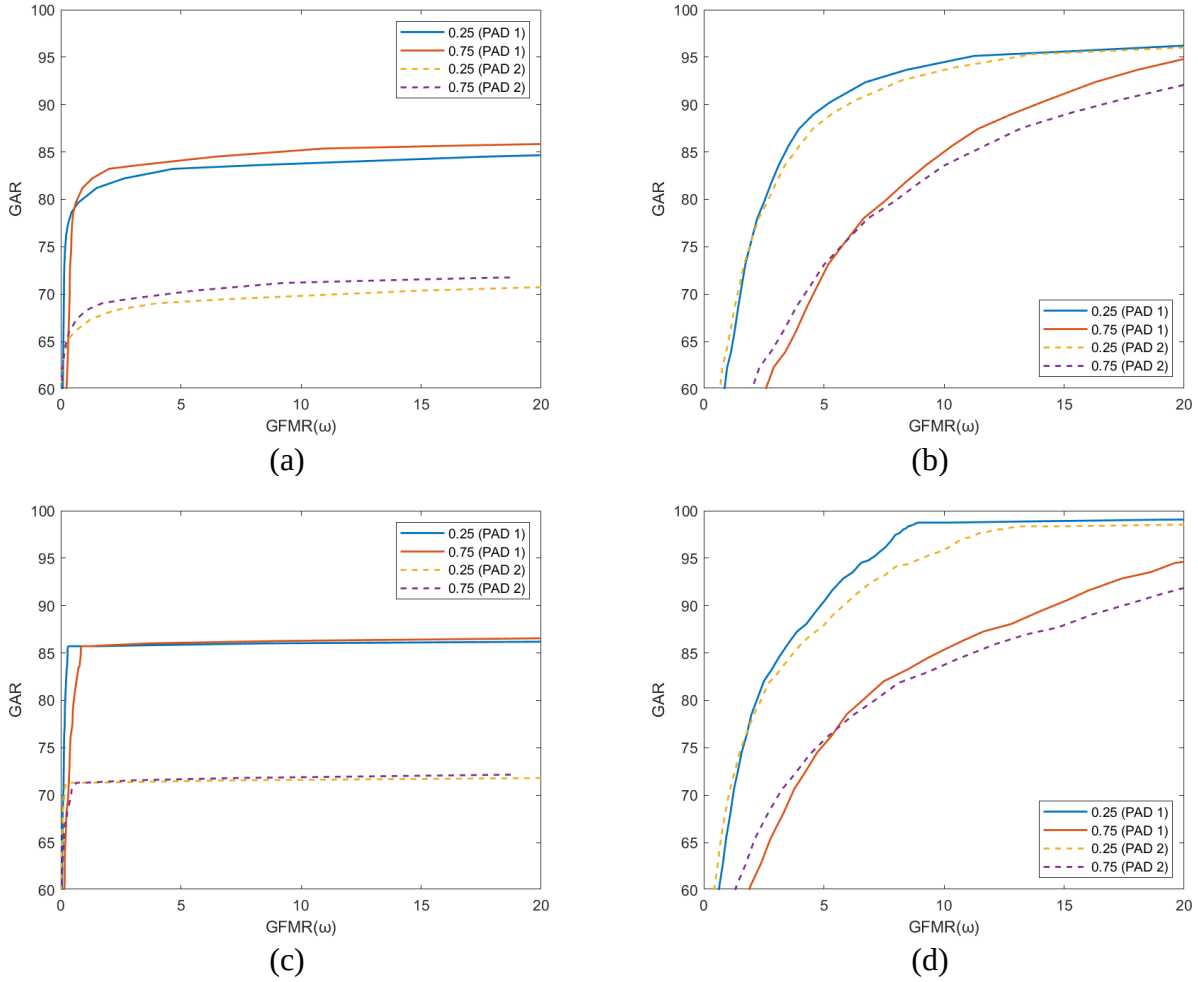


Figure 4.15: DigitalPersona dataset from LivDet 2019. GROC comparison between the top two LivDet 2019 winners, embedded with Bozorth3 (a,b) and Verifinger 12 (c,d) matching system, when the presentation attacks probability w ∈{0.25, 0.75}. Both operational points are reported for each matcher..

4.3.4.4 Application tips: the Global EER analysis

In the previous Sections, we showed how the Bio-WISE output, namely, the integrated system's metrics, allowed us *qualitatively* compare several PAD and matchers. In order to make a choice, the meta-designer must consider several possible probabilities of the occurrence of a presentation attack. Nevertheless, the correct ratio of presentation attacks and impostors in the negative class union can sometimes be unknown at design time. Therefore it may be challenging to evaluate which is the most suitable solution for a given application. To give a more practical instrument in deciding *quantitatively* which operational points the given embedding is worthy of being implemented or not, we considered a well-established criterion: the Equal Error Rate (EER). Since we introduced the GFMR metric, we can simplify the pseudo-ternary classification problem of integrated system so that it suits the binary nature of the verification systems, and therefore define the so-called Global-EER as follows:

$$GEER(\tau^*) = \frac{(GFMR(\tau^*) + (1 - GAR(\tau^*)))}{2} \tag{4.19}$$

where τ^* is the optimal threshold which ensures the minor difference between the GFMR and the complementary of the GAR.

By plotting the GEER at different values and the EER of the verification system alone, we obtain a very informative graph where the point at which the PAD begins to improve the performance of the integrated system can be individuated effortlessly. For instance, let us consider the GEER of the usual integrated system working at $APCER_{01}$ and $BPCER_{01}$ operational points, for all LivDet 2019 datasets and matchers examined (Figure 4.16). We immediately notice that one or both the GEER curves (solid lines) can intersect the EER one (dashed line) at a precise point, the so-called $GEER^*$. If we now consider w^*, the value of w corresponding to the $GEER^*$ point, and $\widehat{w}$ the probability of attack estimated a priori for a specific application, and the following condition occurs:

$$w^* \leq \widehat{w} \tag{4.20}$$

then the designer's choice will fall on the integrated system. In other words, the $GEER^*$ value can be exploited as a parameter for choosing the best solution between the two approaches. Accordingly, the designer now has a concrete and appropriate instrument to investigate the feasibility of such integration *quantitatively*. For example, if we consider the GreenBit sensor, with Verifinger 12 equipped and the PAD working at the $APCER_{01}$ (Fig. 4.16b), the w^* value corresponding to the intersection point of the GEER and EER curves is approximately equal to 0.20. Therefore, if the

estimated attack probability should prove to be lower, the PAD can be turned off/replaced or its operational point changed.

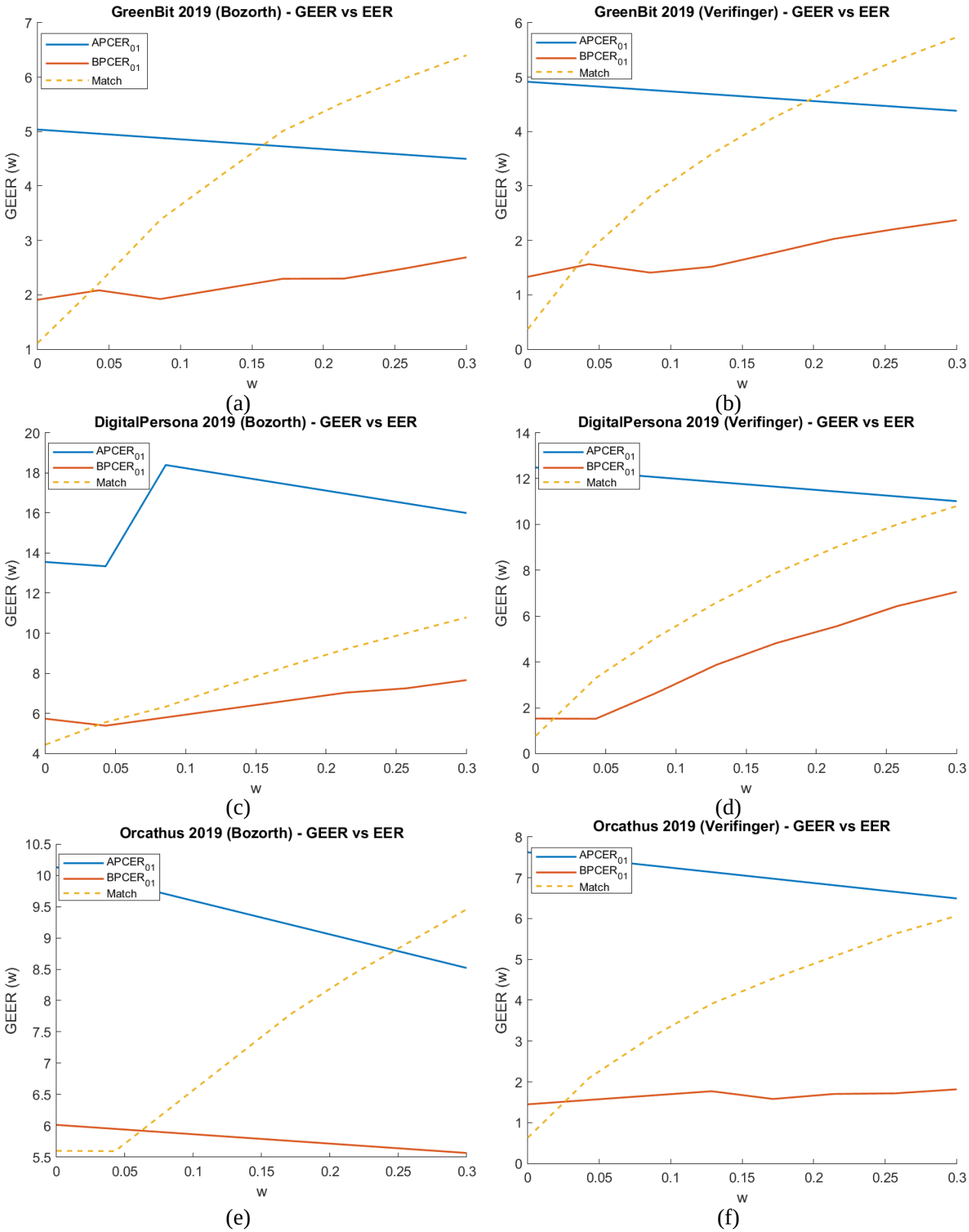


Figure 4.16: LivDet 2019 datasets: GEER trend when the presentation attacks probability $w \in [0.00, 0.30]$. Both matchers are reported: Bozorth3 (left column) and Verifinger 12 (right column). The intersection point of the integrated system curves (solid orange and blue) with the individual system one (dashed yellow) is called $GEER*$. The employed PAD is the "PADUnkFv" algorithm.

4.4 Discussion

Based on our experience gained thanks to the organization of the International Fingerprint Liveness Detection Competition (LivDet) [4], which allows us to make the point about the current performance of presentation attacks detectors, the present work was aimed to make a step ahead, by investigating the integration of PAD and verification system jointly. As a matter of fact, the design of such a system (in general, any biometric verification system) cannot ignore the vulnerability to presentation attacks. In other words, the ability to detect presentation attacks must be integrated into the fingerprint verification system from its first design steps. However, the literature lacked an effective instrument to evaluate the impact of such embedding.

To solve this issue, we developed a model able to predict the performance of the sequential combination of a presentation attacks detector and a fingerprint verification system. Worth remarking, the sequential fusion is one among several possibilities, but it is also the simplest and widespread one. Reported experiments on the LivDet 2017-2019 datasets showed the validity of the proposed model. This allows to appreciate directly the connection between the current PAD performance and matching system. From our investigation, we derived the following guidelines for the designer:

- The w parameter gives a precise picture of the integrated system's performance from the point of view of the probability of a spoofing attack compared to zero-effort attacks. As a result, the designer may adjust w's value at different points to evaluate several security situations. Through the GEER parameter, a critical value of w can be found in which the performance of the integrated system begins to improve concerning that of the verification system alone. If it is out of design constraints, the PAD can be replaced or turned off.

- Setting the operational point of the PAD (parameter in Figure 4.1) allows for observing how the system rates (GAR, FMR and IAPMR) vary. If a too-high GAR loss is registered, the liveness threshold can be relaxed until obtaining a satisfactory result in terms of attacks rejected and genuine accepted.

- By comparing the ROCs of different available systems, Bio-WISE allows for assessing worst-case and best-case scenarios and setting the most suitable PAD accordingly. Our experiments found critical issues with gelatin-like materials and small surface-based sensors (Sections 4.3.4.1 – 4.3.4.2).

On the other hand, the following limitations were pointed out:

- Estimations of the individual ROCs, namely, that of fingerprint presentation attacks detector and verification system, must be accurate if we are interested in the explicit prediction of the fusion performance. This is necessary to reduce the discrepancies between expected and actual performance values.

- The proposed model works only for the sequential combination of PAD and verification systems. Other approaches, like the score-level fusion-based ones, need a specific model. To confirm this, we checked the consistency of the simulator in predicting the performance of integrated system algorithms proposed in LivDet2019, which were not based on sequential fusion. Unfortunately, the obtained estimation error did not allow a reliable adoption of the model. As was expected mainly, this simulator has no general application.

Despite the limitations above, it should be remarked that the prediction of the error rates can be done a priori, that is, before implementing the system on overall. This property can be used, for example, in the meta-design process. Given two possible individual ROC curves, Bio-WISE simulates the performance achievable by their sequential fusion. Showing that the possible performance of the system is around an acceptable range could be enough during the meta-design process. This may lead to a specific performance-oriented approach to the design of an intrinsically secure fingerprint verification system.

5. Bio-WISE upgrade: the *trade-off*

The evaluations carried out so far highlight that the most evident problem of integrating fingerprint PAD and verification system is the loss of genuine acceptance rate (GAR). We presented several insights for the designer to take full advantage of our simulator. There is, nevertheless, a limit. Using the current version of BIO-WISE, it is difficult to appreciate for which PAD's operational points the overall GAR degradation can be still acceptable, with the advantage of handling presentation attacks. The key question is: how much GAR loss can we accept, having at the same time the ability to detect PAs? A novel instrument is needed to answer appropriately, especially considering PADs with different characteristics.

In this Section, we introduce a formal definition of "trade-off", a term that is used when referring to "a balancing of factors all of which are not attainable at the same time"[8]. We carried out a new set of simulations using LivDet 2017 and 2019 data sets, specifically oriented to derive, from the proposed formulation of trade-off, the extent to which the PAD can be integrated without significantly degrading the whole performance, and accordingly, draw the main guidelines for this purpose.

5.1 Performance "*trade-off*": a formal definition

Eqs. 4.15 are not sufficiently expressive for evaluating whether the loss of accepted genuines introduced by the PAD embedding can be kept within a given tolerance range. For this purpose, we introduce a novel parameter, called trade-off, defined as the ratio between the fraction of attackers and genuines accepted for a given match threshold value s_M^*. Since the attackers can be classified into two groups, impostors (zero-effort) and presentation attacks, we have two trade-off values:

$$T_{ZE}^{M} = \frac{FMR(M)}{GAR(M)} \tag{5.1}$$

$$T_{PA}^{M} = \frac{IAPMR(M)}{GAR(M)} \tag{5.2}$$

Where the abbreviations ZE and PA stand respectively for "zero-effort" and "presentation attack". As it can be easily verified, the formal definition above quantifies the "balance" expressed in the definition above. Thus it is reasonable to refer to Eqs. 5.1-5.2 as representatives of the term "trade-

[8] https://www.merriam-webster.com/dictionary/trade-off

off" when a matcher must deal with genuine users and attacks "at the same time". Due to the cumulative nature of the error curves and since $IAPMR \geq FMR$, the relation $T_{ZE} \leq T_{PA}$ is always valid, whatever the matcher threshold value found.

Additionally, these metrics can be successfully employed to assess the worst-case performance scenarios without using a PAD, as, by definition, the verification system is unable to counter a presentation attack. Since we are interested in evaluating the improvement achievable by a sequential integrated system, we may express the relative trade-off values by recalling Eqs. 4.15:

$$T_{ZE}^{S} = \frac{FMR_{Seq}}{GAR_{Seq}} = \frac{FMR(M) \cdot (1 - BPCER(F))}{GAR(M) \cdot (1 - BPCER(F))} = T_{ZE}^{M} \tag{5.3}$$

$$T_{PA}^{S} = \frac{IAPMR_{Seq}}{GAR_{Seq}} = \frac{IAPMR(M) \cdot APCER(F)}{GAR(M) \cdot (1 - BPCER(F))} = T_{PA}^{M} \cdot \tau_F \tag{5.4}$$

From these formulations, we can mainly highlight the following aspects:

- The trade-off values relating to zero-effort attacks are independent of the liveness threshold. In other words, the original relationship between FMR and GAR cannot be changed by any PAD.
- The performance ratio, denoted as τ_F, is always less than one since $APCER(F) \leq 1 - BPCER(F)$ for any liveness threshold.
- The PAD inclusion reduces the maximum error obtainable by the verification system alone, namely T_{PA}^{M} in proportion to the τ_F parameter. For the same liveness operating point, the more efficient the liveness detector is, the better the improvement will be.

To further study the role of the trade-off in the systems embedding, we focus on determining whether an operational point of the liveness detector exists, such as to keep the loss of GAR within a specific tolerance margin.

For this purpose, since T_{PA}^{S} varies according to the performance of the liveness detector, we must firstly set a reference operational point of the matcher within which the system, without the PAD, should work. This process will be detailed in the next Section.

5.1.2 A case study: the Equal Error Rate (EER)

We report here an example of a case study obtained by selecting, for the sake of simplicity, the Equal Error Rate (EER), which can be considered the matcher operational point par excellence. However, our findings can be extended to any other operational point. In this instance, the Eqs. 5.1-5.2 assume the following constant values:

$$T_{ZE}^{EER} = \frac{EER}{1 - EER} \tag{5.5}$$

$$T_{PA}^{EER} = \frac{EER + \Delta}{1 - EER} = T_{ZE}^{EER} \cdot \left(1 + \frac{\Delta}{EER}\right) = T_{ZE}^{EER} \cdot (1 + \Delta_{EER}) \tag{5.6}$$

Where the term Δ_{EER} in Eq. 5.6 expresses the fraction deviation (also representable in percentage) from the reference trade-off T_{ZE}^{EER} and depends on the relative performance difference Δ between the percentage of impostors (FMR) and presentation attacks accepted ($IAPMR$) at the EER. It is worth remarking that these quantities are known from the ROC of the verification system. Similarly, we can easily define from Eq. 5.4 the trade-off for presentation attacks relating to the serial system, as it is the only one subject to the PAD influence:

$$T_{PA}^{S,EER} = T_{PA}^{EER} \cdot \tau_F = T_{ZE}^{EER} \cdot (1 + \Delta_{EER}) \cdot \tau_F \tag{5.7}$$

In order to show how the trade-off values can be exploited to select the most appropriate PAD operating point, we provide in Figure 5.1 a toy example representing a possible trend of the trade-off curves defined by Eqs. 5.5-5.8 when plotted against the GAR of the sequential system. We remember that our model (Eqs. 4.15) may predict the integrated system's indices without actually implementing it overall.

First of all, we observe that the $T_{PA}^{S,EER}$ curve (blue line) is included within the operational points of $zero - APCER$ of the PAD (τ_F = 0) and the first liveness threshold value for which τ_F= 1. At this point, the serial system equals the matcher's performance in detecting spoofs, thus cancelling all PAD advantages. This means that, through this curve, we can precisely define the GAR loss associated with each operational point of the liveness detector. Among these, what is the working point that may guarantee the most appropriate balance? Ideally, the best possible compromise would allow keeping the performance of the integrated system stable on zero-effort and at the same time improve that relating to PAs.

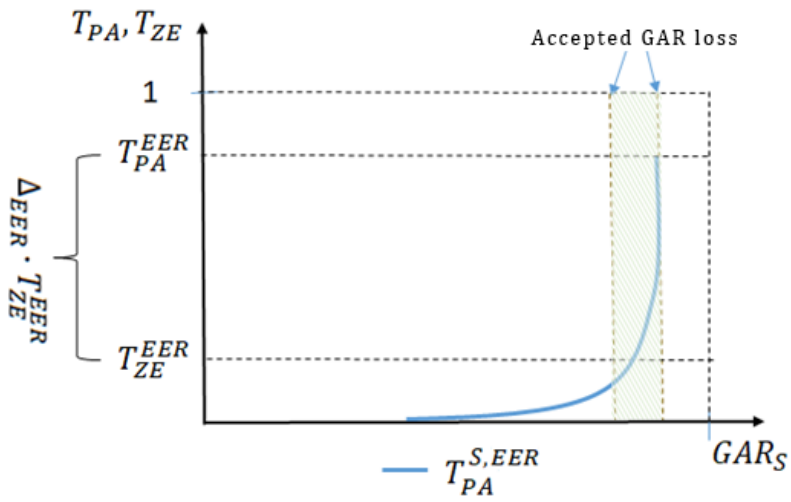


Figure 5.1: Toy graph displaying the relationship between T_{ZE}^{EER}, T_{PA}^{EER} and $T_{PA}^{S,EER}$ in an integrated system.

In the graph, this point corresponds to the intersection of the $T_{PA}^{S,EER}$ curve with the T_{ZE}^{EER} straight line, in which the ratio between impostors/genuine is equivalent to the false/genuine one, namely the integrated system detects fakes with the same ”efficiency” with which the matching system alone blocks the impostors. As previously stated, this results in a GAR loss, which is proportional to the performance of the liveness detector. Accordingly, the tolerance margin, within which to accept the genuines' loss to improve the fakes' detection, can be defined as follows:

$$\rho = GAR(EER) - GAR_{adm} \tag{5.7}$$

where GAR_{adm} is the minimum admissible GAR of the integrated system, which is still compatible with the simulated scenario's constraints. The higher it is, the greater is the tolerance. In this case, ρ indicates the maximum percentage deviation from the nominal GAR value of the matcher at the EER. In our graphs, it is represented by the green area. Once the region has been delimited, we can derive the following guideline from Fig. 5.1: if the accepted GAR loss is on the left or coincident with the point at which the blue curve assumes the value T_{ZE}^{EER} , the most advisable decision is to set the working point for the PAD at that value, keeping the loss of GAR within the fixed range.

On the other hand, it is possible to evaluate any intermediate point that generates a satisfactory advance compared to the verification system's case. Whether such a point does not exist, the PAD under consideration does not fit the scenario's constraints. Finally, the example also suggests that the trade-off on presentation attacks can be made even better than the T_{ZE}^{EER} value, but with significant attention to the fact that the GAR of the sequential system degrades rapidly. Conversely, more fakes are accepted by increasing the percentage of genuine accepted relative to the intersection point. In

summary, once the matcher operational point has been set, our trade-off definition allows to accurately assess under which conditions a presentation attack detector can be integrated without significantly degrading the overall performance in terms of GAR. The following section shows how to apply the outlined guidelines to a real-case study.

5.2 Experimental analysis

5.2.1 Datasets and protocol

The proposed experimental analysis was performed on LivDet 2017, and LivDet 2019 datasets [4]. We analyze all the algorithms submitted to both competitions. Nevertheless, for the sake of space, we report only the results obtained by the top-two winners of LivDet 2019 on the datasets of the same competition (Table 4.1), since their behaviour allows us to summarize that of the other algorithms and cover diverse simulation scenarios. Then, for each dataset, we proceeded as follows: (1) We computed the liveness scores using the two winning algorithms submitted to the LivDet 2019 competition, namely "*PADUnkFv*" and "*JLW LivDet*". Both PADs can be considered amongst the best at the state of the art; (2) We computed the match score using the standard NIST Bozorth3 and the top-level VeriFinger 12 matcher. (3) We derived individual acceptance rates for the matching system, namely GAR, FMR and $IAPMR$, and the error rates for the liveness detector, that is, $BPCER$ and $APCER$, and subsequently, we applied Eqs. 4.15 for computing the acceptance rates of the integrated system. (4) We computed the trade-off values by setting the operating point of the matcher at EER (Eqs. 5.5-5.7). (5) We applied the rules defined in Section 5.1 to define the tolerance margin and consequently set the best trade-off among error rates. This analysis demonstrated that our novel instrument may be employed not only in the meta-design process to determine the optimum PAD operating point, but also as a comparator of current PAD technology when applied to a specific matcher and sensor combination.

5.2.2 Results

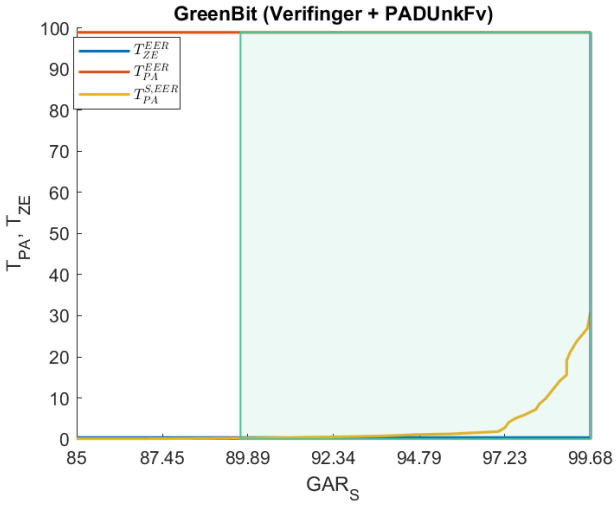
In order to guarantee a correct evaluation of the data and graphs, we first report in Table 5.1 the values of T_{ZE}, T_{PA} and GAR calculated at the EER working point of the matcher for the analyzed sensors. The significant difference between the two trade-off values of zero-effort and presentation attaks testify the danger of spoofing if not correctly contrasted. This is particularly apparent for the Verifinger 12 matcher, which although it provides a benefit to zero-effort attacks detection, presents a much higher T_{PA}^{EER} than Bozorth3 and consequently, it is more vulnerable to presentation attacks.

Therefore, it is necessary to integrate a PAD with the verification system and evaluate the impact on the overall performance to mitigate such risk. Let us consider, for instance, the GreenBit sensor equipped with the matcher VeriFinger and the best detector of LivDet 2019, namely "*PADUnkFv*" (Figure 5.2a). The last value indicated in the x-axis corresponds to the GAR value at the EER. The acceptance area is obtained by setting $\rho = 10\%$. To evaluate the PAD's effectiveness in detecting spoofs, the T_{PA}^{S} (yellow curve) curve should be observed. We note that this curve crosses the straight line T_{ZE}, for a value of $GAR \simeq 92\%$. It means that to bring the liveness detection rate (trade-off on PAs) to the same level of the verification system's accuracy on impostors (trade-off on zero-effort attacks), we should accept a loss of GAR of approximately eight per cent. The crossing point is located within the green area of tolerance, therefore this could be a case of a feasible integrated system, as it can block presentation attacks with high efficiency, keeping the associated GAR loss within the performance constraints.

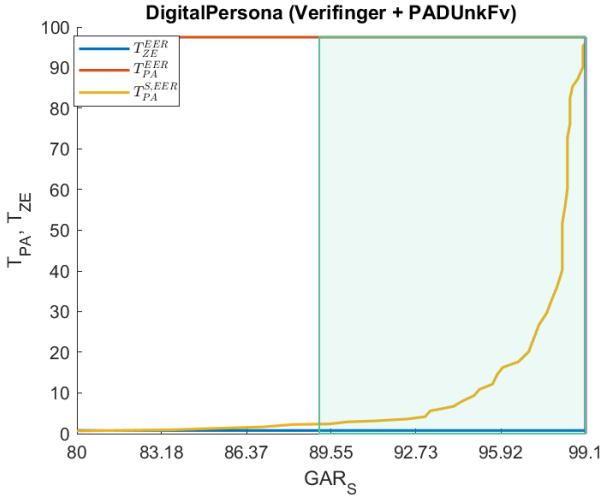
Another example of a suitable embedded system is depicted in Figure 5.2c, obtained by applying the same configuration of PAD/matcher to the Orcanthus sensor. However, it is important to point out that the intersection point is only a possible choice. As a matter of fact, our simulator allows to clearly view the integrated system's behaviour for each PAD operational point and, accordingly, choose the one that best suits the final application context. In both cases (Fig. 5.2{a,c}), the T_{PA}^{S} curve decreases rapidly at first and then more slowly until it crosses the T_{ZE} line; therefore, we could select an intermediate point shortly before the gradient becomes too small, achieving a good compromise between rejected PAs and GAR loss. This also applies when the crossing point is not located within the green area. Figure 5.2b shows a case of this kind, related to the DigitalPersona sensor. Here, the high performance guaranteed by the matcher VeriFinger 12 ($EER < 1\%$) generates a not practicable trade-off point due to the high loss of accepted genuines ($GAR \simeq 83\%$). Nevertheless, the simulation shows us that it is still possible to consistently improve the detection of fakes of over 90 % than in the case of the recognition system alone, by choosing, for instance, the point corresponding to the maximum accepted loss value as the PAD's threshold or any other value within the green area.

Table 5.1: GAR and trade-off values (in percentage) at the EER operational point for Greenbit, DigitalPersona and Orcanthus sensors equipped with Bozorth3 and Verifinger 12 matchers.

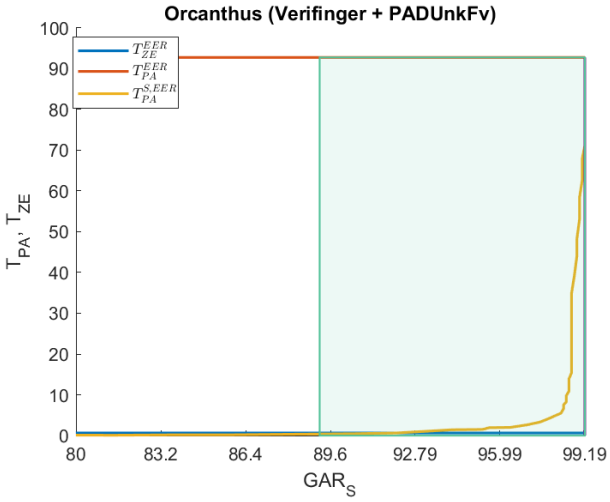
	GreenBit			**DigitalPersona**			**Orcanthus**		
	GAR	T_{ZE}	T_{PA}	GAR	T_{ZE}	T_{PA}	GAR	T_{ZE}	T_{PA}
Bozorth3	98.77	1.13	69.43	94.73	4.66	58.48	95.30	5.93	52.79
Verifinger12	99.68	0.37	98.92	99.10	0.78	97.49	99.20	0.64	92.72



(a)



(b)



(c)

Figure 5.2: $T_{PA}^{S,EER}$ trend for GreenBit (a), DigitalPersona (b) and Orcanthus (c) sensors equipped with Verifinger 12 matching system and "PADUnkFv" liveness detector. The tolerance margin ρ (green area) is set to 10%.

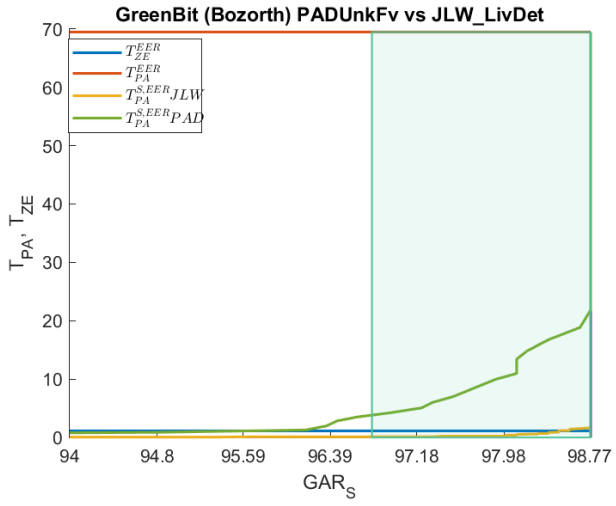
Another advantage of our trade-off definition is the ability to compare several PADs simultaneously, study their behaviour and consequently choose the one that provides better performance. For this purpose, we show the comparison between the two PADs under examination when they are integrated with the Bozorth3 matcher on all the investigated sensors (Figure 5.3). This analysis shows three different scenarios, exemplified by each sensor:

- Figure 5.3a: the $PADUnkFV$ algorithm (green curve) does not meet the required specifications since not only its intersection point is outside the acceptance area defined by the tolerance parameter ($\rho = 2\%$) but also has lower accuracy under the same liveness threshold. On the other hand, the JLW LivDet algorithm (yellow curve) fits the GreenBit characteristics perfectly, achieving a trade-off on PAs comparable to the trade-off on ZE attacks with only 1% loss of GAR.

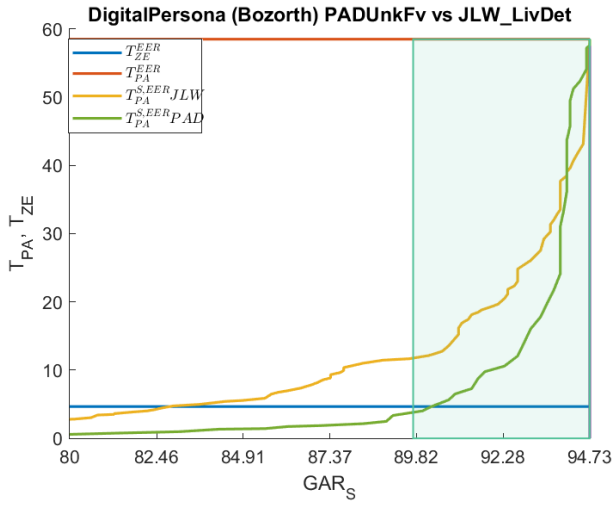
- Figure 5.3b: in this case, the JLW LivDet algorithm (yellow curve) does not match the needed parameters. The tolerance area is defined setting $\rho = 5\%$ and the *PADUnkFV* , albeilt borderline, respects the performance constraints.

- Figure 5.3c shows instead a situation of equality among the two PADs since the two curves are nearly superimposed. Therefore, both could be valid choices in an application scenario.

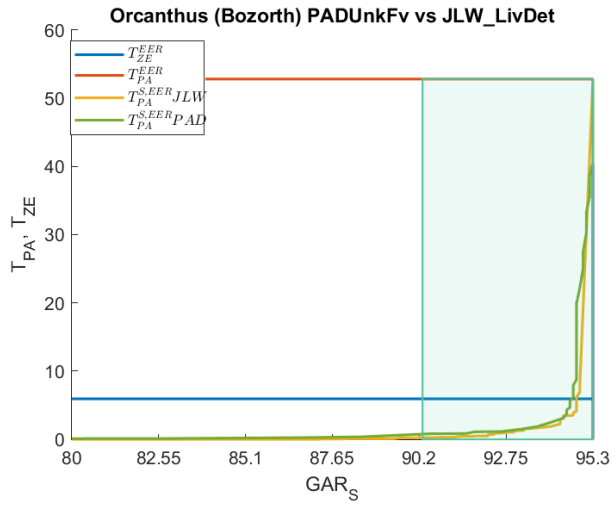
It is worth recalling that we focused only on the top-two winners algorithms of LivDet 2019, nevertheless, the proposed tool can be easily employed to compare several PAD and assess which solution is the most accurate for a given task or simply to evaluate the performance in terms of genuine loss. For this purpose, we present in Figure 5.4 a comparison of eight different PADs submitted to LivDet 2017 embedded with Bozorth3 on DigitalPersona sensor. For the sake of clarity, we did not draw the green area. However, we can immediately notice that the best PAD is the “PAD 7”, with a GAR loss of approximately two percent at the intersection point. This means that the integrated system can work at the EER operational point of the verification system by improving its spoof detection by over 35% with a relatively small cost.



(a)



(b)



(c)

Figure 5.3: Comparison between $T_{PA}^{S,EER}$ curves of the top two PADs of LivDet 2019 integrated with Bozorth3 matcher on the GreenBit (a), DigitalPersona (b) and Orcanthus (c) sensor. The tolerance margin ρ (green area) is set to 2% (a) and 5% (b,c).

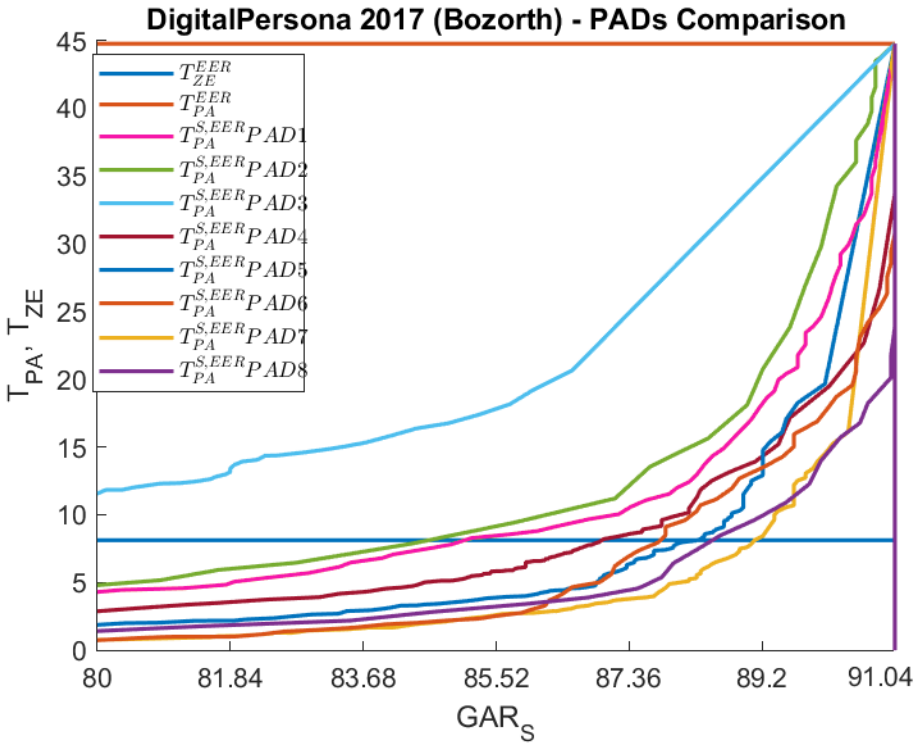


Figure 5.4: Comparison between $T_{PA}^{S,EER}$ of eight different PADs of LivDet 2017 integrated with Bozorth3 matcher on the DigitalPersona sensor.

6. Conclusions

This paper's contribution is the investigation of the probabilistic relationships of error rates when integrating a fingerprint verification system with a presentation attacks detector. Main question was if it is possible to simulate or predict the performance of a fingerprint verification system when sequentially integrated with a PAD. This question was raised by the fact that the liveness detection ability is explicitly required for current personal verification systems, but, on the other side, the performance of PADs does not seem to still fit the high performance requirements which allow to meet the constraints on FMR and GAR. More, was it possible to explicit the probabilistic relationships between acceptances and rejections for the two modules, in order to evaluate if and how they can be sequentially combined, at which operational points they can work? With the goal of answering to the above questions, this paper's main finding is a theoretical model, named Bio-WISE, able to predict the overall system performance from the individual ROC curves of PAD and verification systems, thus simplifying the process of setting the best trade-off among error rates. The proposed model has shown to be reliable according to results obtained on the LivDet data sets. In our opinion, it can be considered as a starting point to approach the design of a fully integrated fingerprint verification from the point of view of the individual performances at hand.

It is also possible to hypothesize appropriate ROCs and make simulations to assess if the fusion performance can be acceptable. The reported results suggest that integrating a PAD into a fingerprint verification system is suitable if the operating point is chosen carefully and the probability of an attack is small but non-zero. This is particularly evident in the GEER analysis shown in Figure 4.16.On the other hand, we pointed out some drawbacks which limit the potential of this model. For instance, due to the difficulty in providing the integrated system's evaluation on multiple operational points, only two PAD's operational points were set. In particular, it was not possible to quantify the balance between the addition of PAD ability and the loss of accuracy on genuine users. Thus, we defined a new measurement, called "trade-off" to investigate every PAD working point. Our trade-off definition allowed to select the most appropriate PAD setting from a theoretical viewpoint. We showed the practical use of such achievement.

To sum up, the literature previously lacked instruments such as BIO-WISE. This and the trade-off presented here can be considered a first pioneering step forward. However, other integration approaches are possible. With BIO-WISE, we wanted to introduce a novel pathway worthy to meet the research community's interest.

Bibliography

[1] Sousedik, C., & Busch, C. (2014). Presentation attack detection methods for fingerprint recognition systems: a survey. *Iet Biometrics*, *3*(4), 219-233.

[2] Marasco, E., & Ross, A. (2014). A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)*, *47*(2), 1-36.

[3] Marcel, S., Nixon, M. S., & Li, S. Z. (2014). *Handbook of biometric anti-spoofing* (Vol. 1, p. 96). London: Springer.

[4] Micheletto, M., Orrù, G., Casula, R., Yambay, D., Marcialis, G. L., & Schuckers, S. C. (2022). Review of the Fingerprint Liveness Detection (LivDet) competition series: from 2009 to 2021. *arXiv preprint arXiv:2202.07259*.

[5] Abhyankar, A., & Schuckers, S. (2009). Integrating a wavelet based perspiration liveness check with fingerprint recognition. *Pattern Recognition*, *42*(3), 452-464.

[6] Marasco, E., Ding, Y., & Ross, A. (2012). Combining match scores with liveness values in a fingerprint verification system. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (pp. 418-425). IEEE.

[7] Rattani, A., Poh, N., & Ross, A. (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)* (pp. 37-42). IEEE.

[8] Busch, C. (2019). Standards for biometric presentation attack detection. In *Handbook of Biometric Anti-Spoofing* (pp. 503-514). Springer, Cham.

[9] Chingovska, I., Anjos, A., & Marcel, S. (2013). Anti-spoofing in action: joint operation with a verification system. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 98-104).

[10] Marcel, S., Nixon, M. S., & Li, S. Z. (2014). *Handbook of biometric anti-spoofing* (Vol. 1, p. 96). London: Springer.

[11] Marasco, E., Johnson, P., Sansone, C., & Schuckers, S. (2011, June). Increase the security of multibiometric systems by incorporating a spoofing detection algorithm in the fusion mechanism. In *International Workshop on Multiple Classifier Systems* (pp. 309-318). Springer, Berlin, Heidelberg.

[12] Rattani, A., & Poh, N. (2013, June). Biometric system design under zero and non-zero effort attacks. In *2013 International Conference on Biometrics (ICB)* (pp. 1-8). IEEE.

[13] Poh, N., Wong, R., & Marcialis, G. L. (2014). Toward an attack-sensitive tamper-resistant biometric recognition with a symmetric matcher: A fingerprint case study. In *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)* (pp. 175-180). IEEE.

[14] Chingovska, I., Mohammadi, A., Anjos, A., & Marcel, S. (2019). Evaluation methodologies for biometric presentation attack detection. In *handbook of biometric anti-spoofing* (pp. 457-480). Springer, Cham.

[15] Micheletto, M., Marcialis, G. L., Orrù, G., & Roli, F. (2021). Fingerprint recognition with embedded presentation attacks detection: are we ready?. *IEEE Transactions on Information Forensics and Security*, *16*, 5338-5351.

[16] Jain, A. K., Flynn, P., & Ross, A. A. (Eds.). (2007). *Handbook of biometrics*. Springer Science & Business Media.

[17] Jain, A., Bolle, R., & Pankanti, S. (Eds.). (1999). *Biometrics: personal identification in networked society* (Vol. 479). Springer Science & Business Media.

[18] Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, *14*(1), 4-20.

[19] Galbally Herrero, J., Fierrez, J., & Ortega-García, J. (2007). Vulnerabilities in biometric systems: Attacks and recent advances in liveness detection. -.

[20] Ekenel, H. K., & Stiefelhagen, R. (2009). Why is facial occlusion a challenging problem?. In *International conference on biometrics* (pp. 299-308). Springer, Berlin, Heidelberg.

[21] Hassaballah, M., & Aly, S. (2015). Face recognition: challenges, achievements and future directions. *IET Computer Vision*, *9*(4), 614-626.

[22] Ross, A., & Jain, A. (2004, May). Biometric sensor interoperability: A case study in fingerprints. In *International Workshop on Biometric Authentication* (pp. 134-145). Springer, Berlin, Heidelberg.

[23] Willis, D., & Lee, M. (1998). Six biometric devices point the finger at security. *Computers & Security*, *5*(17), 410-411.

[24] ISO/IEC 30107-3:2017(en). (2017). Information Technology-Biometric Presentation Attack Detection-Part 3: Testing and Reporting. Available online: https://www.iso.org/obp/ui/#iso:std:iso-iec:30107:-3:ed-1:v1:en (accessed on 30 October 2022).

[25] Li, S. Z. (2009). *Encyclopedia of Biometrics: I-Z* (Vol. 2). Springer Science & Business Media.

[26] Poh, N., & Bengio, S. (2006). Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication. *Pattern Recognition*, *39*(2), 223-233.

[27] Maltoni, D., Maio, D., Jain, A. K., & Prabhakar, S. (2009). *Handbook of fingerprint recognition*. Springer Science & Business Media.

[28] Jain, A. K., Flynn, P., & Ross, A. A. (Eds.). (2007). *Handbook of biometrics*. Springer Science & Business Media.

[29] Dyre, S., & Sumathi, C. P. (2016). *A survey on various approaches to fingerprint matching for personal verification and identification",* Int. j. comput. sci. eng. surv., vol. 7, no. 4, pp. 01-17.

[30] Nandakumar, K., & Jain, A. K. (2004). Local Correlation-based Fingerprint Matching. In *ICVGIP* (pp. 503-508).

[31] Ross, A., Reisman, J., & Jain, A. (2002). Fingerprint matching using feature space correlation. In *International Workshop on Biometric Authentication* (pp. 48-57). Springer, Berlin, Heidelberg.

[32] Gutierrez, P. D., Lastra, M., Herrera, F., & Benítez, J. M. (2013). A high performance fingerprint matching system for large databases based on GPU. *IEEE Transactions on Information Forensics and Security*, *9*(1), 62-71.

[33] Qi, J., & Wang, Y. (2005). A robust fingerprint matching method. *Pattern recognition*, *38*(10), 1665-1671.

[34] Marana, A. N., & Jain, A. K. (2005, October). Ridge-based fingerprint matching using hough transform. In *XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'05)* (pp. 112-119). IEEE.

[35] Masmoudi, A. D., & Masmoudi, D. S. (2010). Implementation of a fingerprint recognition system using LBP descriptor. *Journal of Testing and Evaluation*, *38*(3), 369-382.

[36] Nanni, L., & Lumini, A. (2009). Descriptors for image-based fingerprint matchers. *Expert Systems with Applications*, *36*(10), 12414-12422.

[37] Matsumoto, T., Matsumoto, H., Yamada, K., & Hoshino, S. (2002, April). Impact of artificial" gummy" fingers on fingerprint systems. In *Optical Security and Counterfeit Deterrence Techniques IV* (Vol. 4677, pp. 275-289). SPIE.

[38] Arora, S. S., Jain, A. K., & Paulter, N. G. (2017). Gold fingers: 3D targets for evaluating capacitive readers. *IEEE transactions on information forensics and security*, *12*(9), 2067-2077.

[39] Marasco, E., & Ross, A. (2014). A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)*, *47*(2), 1-36.

[40] Marasco, E., Cando, S., Tang, L., Ghiani, L., & Marcialis, G. L. (2018, November). A look at non-cooperative presentation attacks in fingerprint systems. In *2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-6). IEEE.

[41] Casula, R., Micheletto, M., Orrú, G., Marcialis, G. L., & Roli, F. (2022). Towards realistic fingerprint presentation attacks: the ScreenSpoof method. *Pattern Recognition Letters*.

[42] Putte, T. V. D., & Keuning, J. (2000). Biometrical fingerprint recognition: don't get your fingers burned. In *Smart Card Research and Advanced Applications* (pp. 289-303). Springer, Boston, MA.

[43] Lapsley, P. D., Lee, J. A., Pare Jr, D. F., & Hoffman, N. (1998). *U.S. Patent No. 5,737,439*. Washington, DC: U.S. Patent and Trademark Office.

[44] Baldisserra, D., Franco, A., Maio, D., & Maltoni, D. (2006, January). Fake fingerprint detection by odor analysis. In *International Conference on Biometrics* (pp. 265-272). Springer, Berlin, Heidelberg.

[45] Biel, L., Pettersson, O., Philipson, L., & Wide, P. (2001). ECG analysis: a new approach in human identification. *IEEE Transactions on Instrumentation and Measurement*, *50*(3), 808-812.

[46] Wang, F., Cheng, J., & Jiang, Y. (2015). Ridge-Slope-Valley Feature for Fingerprint Liveness Detection. In *The Proceedings of the Third International Conference on Communications, Signal Processing, and Systems* (pp. 857-865). Springer, Cham.

[47] Ghiani, L., Hadid, A., Marcialis, G. L., & Roli, F. (2013, September). Fingerprint liveness detection using binarized statistical image features. In *2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS)* (pp. 1-6). IEEE.

[48] Agarwal, S., Rattani, A., & Chowdary, C. R. (2021). A comparative study on handcrafted features v/s deep features for open-set fingerprint liveness detection. *Pattern Recognition Letters*, *147*, 34-40.

[49] Biggio, B., Fumera, G., Marcialis, G. L., & Roli, F. (2016). Statistical meta-analysis of presentation attacks for secure multibiometric systems. *IEEE transactions on pattern analysis and machine intelligence*, *39*(3), 561-575.

[50] Biometric Best Practices: Optimizing Security in Fingerprint (2014), Crossmatch, White Paper, http://www.crossmatch.com.

[51] Nogueira, R. F., de Alencar Lotufo, R., & Machado, R. C. (2016). Fingerprint liveness detection using convolutional neural networks. *IEEE transactions on information forensics and security*, *11*(6), 1206-1213.

[52] Ross, A., & Jain, A. (2003). Information fusion in biometrics. *Pattern recognition letters*, *24*(13), 2115-2125.

[53] Gonzalez-Soler, L. J., Gomez-Barrero, M., Chang, L., Pérez-Suárez, A., & Busch, C. (2021). Fingerprint presentation attack detection based on local features encoding for unknown attacks. *IEEE Access*, *9*, 5806-5820.