applied sciences

MDPI

Article

Mitigating Sensor and Acquisition Method-Dependence of Fingerprint Presentation Attack Detection Systems by Exploiting Data from Multiple Devices

Marco Micheletto, Giulia Orrù, Roberto Casula and Gian Luca Marcialis *

Department of Electrical and Electronic Engineering, University of Cagliari, 09123 Cagliari, Italy
* Correspondence: marcialis@unica.it

Abstract: The problem of interoperability is still open in fingerprint presentation attack detection (PAD) systems. This involves costs for designers and manufacturers who intend to change sensors of personal recognition systems or design multi-sensor systems, because they need to obtain sensor-specific spoofs and retrain the system. The solutions proposed in the state of the art to mitigate the problem still require data from the target sensor and are therefore not exempt from the problem of obtaining new data. In this paper, we provide insights for the design of PAD systems thanks to an overview of an interoperability analysis on modern systems: hand-crafted, deep-learning-based, and hybrid. We investigated realistic use cases to determine the pros and cons of training with data from multiple sensors compared to training with single sensor data, and drafted the main guidelines to follow for deciding the most convenient PAD design technique depending on the intended use of the fingerprint identification/authentication system.

Keywords: fingerprint; presentation attack detection; interoperability

Citation: Micheletto, M.; Orrù, G.; Casula, R.; Marcialis, G.L. Mitigating Sensor and Acquisition Method-Dependence of Fingerprint Presentation Attack Detection Systems by Exploiting Data from Multiple Devices. *Appl. Sci.* **2022**, *12*, 9941. https://doi.org/10.3390/app12199941

Academic Editors: Zhaoqiang Xia, Haixi Zhang and Jun Wu

Received: 30 August 2022
Accepted: 26 September 2022
Published: 2 October 2022

1. Introduction

Biometric technologies are gaining popularity owing to their reliability and ease. Faces, fingerprints, and other biometric traits can identify individuals with a high degree of accuracy. Among these, the fingerprint is the most widespread due to its uniqueness and ease of use. However, artificial fingerprint replicas, also called spoofs or presentation attacks (PAs), can circumvent fingerprint-based personal recognition. Both consensual and non-consensual procedures may be used to replicate a fingerprint. Consensual techniques need the awareness and cooperation of the intended person, whereas nonconsensual approaches are more dangerous and covert and are based on the acquisition of latent fingerprints left unknowingly on reflective or partially reflective surfaces. Given that automated fingerprint identification systems (AFIS) frequently secure sensitive and critical data, it is crucial to equip them with presentation attack detectors (PADs) that help determinine whether the fingerprint acquired by the sensor is genuine (live) or a replica (fake). Although PADs are nowadays very accurate, their development and use still present open problems.

One of these is the so-called lack of interoperability. A PAD trained on data from one sensor works precisely on images of that sensor. However, it can lead to much lower results when used on images of different sensors. This is a frequent problem when a manufacturer or user needs to replace the sensor of an authentication system with a more current and high-performing device. In this case, in fact, it is not possible to use the previous PAD; the system must be retrained, fine-tuned, or updated on new images. Obtaining these new images may not be easy, as it requires multiple acquisitions of real users and fake fingerprints made with different materials and techniques. Although some recently proposed works aim to mitigate the problem of interoperability [1], they require a subset of the target sensor samples. This makes interoperability an open and critical issue for AFIS systems, especially when obtaining target samples at the design stage is difficult or even infeasible.

Another open problem is the lack of ability of modern PADs to recognize PAs obtained with acquisition techniques and/or materials not used during the training phase. Even in this case, to make the PAD robust and able to recognize the greatest number of types of PAs, a designer must predict, obtain, and use them in the training phase. Obtaining PAs made with all known state-of-the-art techniques and materials can be very expensive.

As the organizers of the Fingerprint Liveness Detection Competition (LivDet) (https://livdet.diee.unica.it/, accessed on 19 September 2022) we were able to observe some different PAD training techniques among researchers. In fact, most of the LivDet competitors trained their PADs on single-sensor data. This training procedure is recommended and rewarded for the purposes of the competition because it allows a precise reading of the results and associates cause and effect in the proposed challenges. However, other competitors used all the data of the training sets from different acquisition sensors simultaneously, and others on additional data compared to the training sets provided. These procedures proved successful in the latest edition of the competition, LivDet 2021, obtaining the best results in terms of presentation attack detection, even on the "advanced" never-seen-before attacks perpetrated by a new semi-consensual acquisition technique called ScreenSpoof [2]. Moreover, these large-trained algorithms have been shown to be particularly robust across the sensors of the competition. Based on this evidence, we wondered to which extent training on data acquired with multiple sensors helps detect artificial replicas. Is this procedure truly beneficial in solving the interoperability and/or generalization problem?

To properly answer, we designed different PADs from scratch and investigated how the performances change by varying the composition of the training set, exploring different interoperability scenarios using LivDet 2013, 2019, and 2021 data sets. As a matter of fact, although the problems of interoperability and generalization are well known at the state of the art [3], this work aims to systematically analyze the various real application contexts in which a PAD can be used. In particular, we represented and examined different combinations between training and use contexts (intra-method and intra-sensor, cross-method, cross-sensor, etc.) and then evaluated which ones can preferably be addressed with a single-sensor training approach, i.e., with a PAD trained on data acquired from a single sensor or with a multi-sensor training approach, that is, a PAD trained on data acquired from different sensors. These experiments allowed us to outline the limits and potentials of single-sensor and multi-sensor training solutions when tested on entirely or partially unknown data.

2. Fingerprint Presentation Attack Detection and Interoperability

2.1. FPAD

The threat of fingerprint spoofs has been known since 1998, the year of publication of the first paper that demonstrated the vulnerability of fingerprint sensors to artificial replicas [4]. A few years later, the first hardware and software countermeasures were proposed [5,6]. The PAD software systems focus on anatomical, physiological, and textural properties and other features that can be extracted and used for matching purposes. As in all fields of pattern recognition, the feature extraction and image classification phases for the detection of presentation attacks have passed from hand-crafted approaches and shallow classifiers to the deep-learning era. Among the hand-crafted extraction methods are local descriptors such as SIFS, BSIF, LBP, and LPQ [7,8]. These approaches use a binary code to characterize each pixel's neighborhood, which is derived by computing the image's convolution with a collection of linear filters and then binarizing the filter responses. The resulting feature vectors are often inputs to shallow classifiers such as SVM [9], but can be input to complex neural networks. Other deep learning techniques are often based on convolutional neural networks (CNN) [10]. They provide extremely high accuracy, but require many training data and a significant amount of time and resources. In recent years, the two types of techniques, hand crafted and deep learning-based, have been increasingly combined to overcome limitations such as the difficulty of modern PADs to generalize and the need for extensive computational resources. One example is Fingerprint Spoof

Buster [11], which uses locally centered and aligned patches using fingerprint minutiae to train a MobileNet-v1 model.

2.2. The Interoperability and Generalization Problem

The interoperability challenge is fundamentally connected to the hardware and software variations between fingerprint acquisition sensors. Optical, solid-state, and ultrasonic sensors are the three most common types of acquisition hardware. Each type of sensor produces a particular distortion in the image linked to a different physical phenomenon, which encodes the valleys and ridges. In addition to the kind of acquisition, scanners can be grouped by image characteristics such as DPI (dot per inch), scan area, geometric precision, etc. The interoperability problem across sensors was recognized in the first instance from the point of view of fingerprint verification. Numerous works over the years have proposed effective solutions to mitigate the drop in accuracy due to the scanner change [12,13]. Recently, this problem has also been addressed in the field of presentation attack detection, proving to be particularly critical. As a matter of fact, many modern solutions still suffer poor generalization performance when tested on data not seen during the training phase, leading to spoof detection error rates up to three times higher [14]. One potential reason for this complexity in generalizing across sensors is that fingerprint images from different sensors possess different textural characteristics (Figure 1). This complexity is exacerbated by the employment of different acquisition procedures and materials between the training and system use phases, which provides an additional degree of freedom to the fingerprint's appearance.

Figure 1. Examples of acquisitions with different sensors: (**a**) Biometrika, (**b**) DigitalPersona, (**c**) Orcanthus, (**d**) GreenBit, (**e**) Dermalog. The characteristics of the images are strictly influenced by the acquisition technology.

In recent years, numerous works have aimed at overcoming these limits. For instance, authors in [1,10,15] mitigated such differences by designing a style-transfer wrapper to add on top of any CNN architecture in order to reduce the performance gap due to cross-sensor evaluations. Their idea is to use a limited number of live fingerprints of the "new" sensor, hereinafter called the *target* sensor, in order to project its style into live and spoof samples coming from the "old" sensor. This new synthetically generated dataset is then employed

to train a liveness detector from scratch, improving the average cross-sensor spoof detection performance by approximately 13%.

Another plausible justification for the problem of interoperability is that the different ways of representing the ridges and valleys of a fingerprint affects the frequencies of the grey-level histogram of the images. Figure 2 supports this idea, where we can observe the mean grayscale histogram for the five sensors investigated in this paper. Starting from this peculiarity, Tuveri et al. [16] reached a significant level of interoperability by shifting sensor-specific feature distributions based on the least squares algorithm. The feature vectors were calculated by using textural algorithms. This solution is optimal since it does not require any additional PAD training. Nevertheless, both live and spoof samples are needed to relocate the feature space of the target sensor into that of the original sensor.

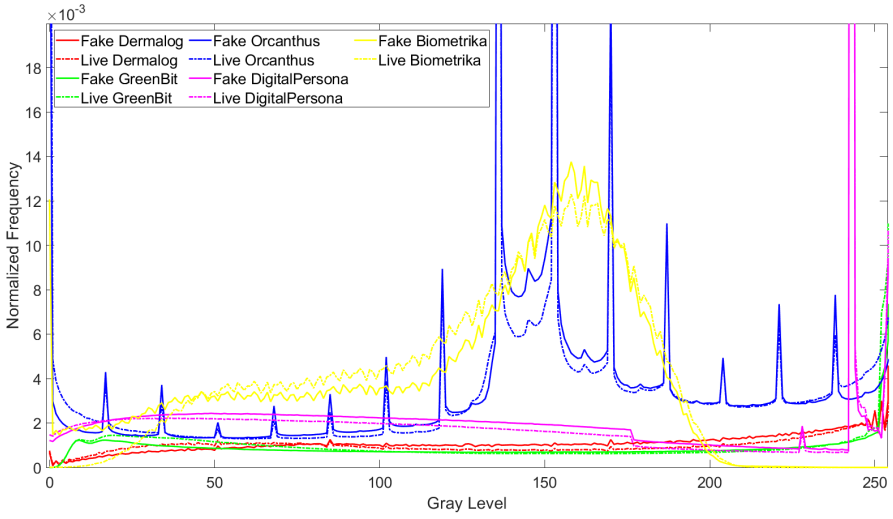


Figure 2. Mean histogram of grayscale for all the five considered datasets.

All these works generally acknowledge the need for a subset of the target sensor's samples to solve the interoperability problem. This inevitably leads to issues from an economic point of view. In fact, effort and money are necessary to collect a spoof dataset; it is essential to find volunteers willing to donate their fingerprints, fabricate fakes of each finger with a substantial number of materials, and finally acquire them via the scanner. Furthermore, based on the employed materials, costs can grow considerably.

The procedure shown in [15] is not affected by these obstacles. On the other hand, it substitutes the original FPAD with a new one tailored to the target sensor. In many applications, this procedure is not feasible. Moreover, this implies that it would be challenging for companies to scale up FPAD tools since each scanner would require its custom detector.

Therefore, with no access to the target sensor data, is it possible to mitigate the sensor dependence on an FPAD system? The goal of this work is to answer this question.

3. Interoperability Scenarios for the Design of Fingerprint PAD

In this paper, we investigated to what extent the interoperability problem afflicts modern PAD systems and whether some solutions such as training on multiple types of data are beneficial or counterproductive. In particular, we want to evaluate whether training on data from different sensors helps the system to generalize or if it allows to it recognize more types of data, but lowers the general performance. For this purpose, we identified the following project scenarios based on the concept of interoperability applied to fingerprint sensors and the PAI fabrication methods:

- Intra-sensor and intra-method: this is the standard and optimal application scenario. The PAD is used to analyze and classify images acquired with the same sensor used in the training phase. The type of PAI used to attack the system is known by the system;

- Intra-sensor and cross-method: this is a standard, but unfavorable application scenario. The PAD is used to analyze and classify images acquired with the same sensor used in the training phase. The type of PAI used to attack the system are unknown to the system. Such a scenario is unpredictable by the designer as new replication methods can be discovered and used after the PAD has been designed and trained;
- Cross-sensor and intra-method: the designer/manufacturer decides, knowing the risks, to use a PAD trained on a sensor on a AFIS consisting of a different acquisition sensor. This choice is not optimal, but is made for economic reasons or in the absence of data to retrain/fine-tune a new PAD. The type of PAI used to attack the system is known by the system, but since the acquisition sensor is different, the resulting images could be very different;
- Cross-sensor and cross-method: as in the previous scenario, the designer uses a PAD trained on a sensor on an AFIS consisting of a different acquisition sensor. Moreover, the type of PAI used to attack the system is unknown to the system.

The first two scenarios exemplify the standard functioning of a PAD and allow us to evaluate the system's robustness to never-seen-before attacks.

On the other hand, cross-sensor scenarios assess the ability of PADs to generalize concerning the scanner change. These are typically non-optimal cases, useful to assess a situation in which a designer cannot easily collect the data with the new sensor or cannot replace the PAD.

Experimental Protocol

To replicate the application scenarios identified in the previous section, we carried out the following experiments:

- Intra-sensor and intra-method: the training set is partially or totally composed of data belonging to the target sensor. The PAIs are created with the same method in the two sets;
- Intra-sensor and cross-method: the training set is partially or totally composed of data belonging to the target sensor. The spoofs of the training set were created with a different method than the test set ones;
- Cross-sensor and intra-method: the training set does not contain data on the target sensor. The PAIs are created with the same method in the two sets;
- Cross-sensor and cross-method: the training set does not contain information on the target sensor. The spoofs of the training set were created with a different method than the test set ones.

For each application scenario, we divided the experimentation into two protocols based on the designer's choice to use a pre-trained model or to train a model from scratch:

- *Pre-trained*: some competitors from the eighth edition of the Fingerprint Liveness Detection Competition have been selected (Table 1). Certain of them used additional data for training, although the use of only the LivDet 2021 training dataset was recommended. This experiment is strongly representative of the current state of the art, but it is not completely controlled, since the implementation details are unknown;
- *Self-trained*: the experiments are fully controlled and the details and training data are known. In particular, (i) two hand-crafted PADs have been implemented consisting of a feature extractor via BSIF and LBP, respectively, followed by a linear SVM classifier and (ii) one deep-learning based PAD. The Spoof Buster method is implemented.

LBP and BSIF allow for obtaining a statistically meaningful representation of the fingerprint data by respectively applying hand-crafted and a fixed set of filters on sub-portions of the image. The feature extraction step is followed by a shallow classifier, a linear SVM. On the other hand, the Spoof Buster method is based on a CNN trained on local patches centered and aligned using minutiae location and orientation.

The performances of the FPADs were evaluated using the ISO metrics [17], in particular, (i) APCER (attack presentation classification error rate), that is, the rate of misclassified

fake fingerprints; (ii) BPCER (bona-fide presentation classification error rate), that is, the rate of misclassified live fingerprints; and (iii) liveness accuracy, the percentages of samples correctly classified by the PAD.

Table 1. Characteristics of the pre-trained PAD LivDet 2021 competitors used for the analysis.

Participant	Algorithm Name	Type	Acronym
Dermalog	LivDet21ColC2	Deep-learning	Col
	LivDet21DobC2	Deep-learning	Dob
Unesp	contreras	Hand-crafted	con
Hangzhou Jinglianwen Tech. Co., Ltd.	JLWLivDetL	Hybrid	JLW
MEGVII (BEIJING) Technology Co., Ltd.	megvii_single	Deep-learning	m_s
	megvii_ensemble	Deep-learning	m_e
University of Applied Sciences Darmstadt	PADUnk	Hand-crafted	PAD
Chosun University	B_ld2	Deep-learning	bld
Anonymous	bb8	Hybrid	bb8
	r2d2	Hybrid	r2d2

4. Results

4.1. Dataset

In general, the cross-sensor analysis can be viewed as two separate cases: (i) all sensors in the evaluation utilize the same sensing technology, and (ii) the sensing mechanisms differ. We explored both cases and, in this respect, we utilized datasets from LivDet 2013, 2019, and 2021 competitions [18] in our experimental analysis. They consist of live and spoof fingerprint images from five different devices, which differ in scan area and sensing technology (Table 2). The spoof images were collected using cooperative and non-cooperative methods based on the LivDet edition.

Table 2. Device characteristics for LivDet 2013, 2019, and 2021 datasets.

Scanner	Model	Resolution [dpi]	Image Size [px]	Format	Type
Biometrika	FX2000	569	315x372	PNG	Optical
Orcanthus	Certis2 Image	500	300xN	PNG	Thermal swipe
DigitalPersona	U.are.U 5160	500	252x324	PNG	Optical
GreenBit	DactyScan84C	500	500x500	BMP	Optical
Dermalog	LF10	500	500x500	PNG	Optical

Furthermore, the materials used to fabricate the PAIs vary across the training and the test sets, as reported in Tables 3 and 4. Due to this lack of knowledge of the spoofing nature, we can simulate a true attack scenario, following the typical fingerprint PAD evaluation protocols.

Table 3. Number of samples for each scanner employed in the training phase.

Training Sets	Live	Latex	RProFast	WoodGlue	Ecoflex	Gelatine
LivDet 2021 GreenBit Training	1250	750	750	-	-	-
LivDet 2021 Dermalog Training	1250	750	750	-	-	-
LivDet 2019 DigitalPersona Training	1000	250	-	250	250	250

Table 4. Number of samples for each scanner employed in the test phase.

Dataset	Test Set					
LivDet 2021	Live	Mix1		BodyDouble	ElmersGlue	
GreenBit CC/SS	2050	820		820	820	
LivDet 2021	Live	GLS20		RFast30		
Dermalog CC/SS	2050	1230		1230		
LivDet 2019	Live	Mix1		Mix2	Liquid Ecoflex	
Orchantus	990	384		308	396	
LivDet 2013	Live	Ecoflex	Gelatine	Latex	Modasil	WoodGlue
Biometrika	1000	200	200	200	200	200

4.2. Pre-Trained Analysis

To evaluate how much the problem of interoperability affects modern PADs, we studied the behavior of a selection of competitors in the latest edition of the LivDet competition. The LivDet 2021 competitors' detectors are typically composed of two models, each trained on a different type of data: a set acquired with the GreenBit sensor and one with the Dermalog sensor, the characteristics of which are reported in Section 4.1. Nevertheless, some competitors claimed to have obtained a single model by training on both sets simultaneously; others added additional data to the training data. These cases will be marked with a single (*) or double (**) asterisk, respectively. The results reported in the Tables 5 and 6 refer to the GreenBit and Dermalog test sets, respectively. Although half of the PADs show a drop in the cross-sensor performance, the other half maintains the same accuracy. We hypothesized that PADs with the same or similar intra-sensor and cross-sensor accuracy were related to models trained on both sensors or additional data. This hypothesis is confirmed by the competitors, who have declared this training approach (marked with the asterisk). The PADs that have this behavior, that is, *LivDet_Col_C2*, *LivDet_DOB_c2*, *megvii*, and *PADUnk*, are also the best detectors of the eighth edition of the competition on the cross-method data, that is, the spoofs acquired with the semi-consensual ScreenSpoof (SS) technique. We therefore wondered if the training technique on multiple types of data could mitigate the problem of interoperability or make the PADs more robust to attacks "never seen before", such as fakes made on new materials or with new acquisition techniques. However, these systems, except for *megvii*, are also the ones that, in the intra-dataset scenario, have a higher APCER and perform worse.

Table 5. Results of LivDet 2021 competitors on test sets acquired with the GreenBit sensor. In particular, GB CC is a test set acquired with the consensual technique, while GB SS is a test set acquired with the ScreenSpoof technique. (*) indicates training on both GreenBit and Dermalog Livdet training sets. (**) indicates training on additional data with respect to LivDet training sets.

Alg.	Trained on GB and Tested on GB CC			Trained on DL and Tested on GB CC			Trained on GB and Tested on GB SS			Trained on DL and Tested on GB SS		
	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]
Col (**)	0.20	29.88	83.61	0.24	23.78	86.92	0.20	24.76	86.41	0.24	21.38	88.23
Dob (**)	0.59	25.41	85.88	0.44	29.84	83.53	0.59	3.25	97.96	0.44	4.72	97.23
con	8.98	3.94	93.77	1.85	80.69	55.14	8.98	26.67	81.37	1.85	94.55	47.58
JLW	2.59	8.21	94.35	0.20	87.76	52.04	2.59	54.11	69.31	0.20	79.76	56.41
m_s (*)	0.29	6.30	96.43	0.29	6.30	96.43	0.29	13.94	92.26	0.29	13.95	92.26
m_e (*)	0.05	2.72	98.49	0.05	2.72	98.49	0.05	13.62	92.55	0.05	13.62	92.55
PAD (*)	1.46	37.20	79.05	1.46	37.20	79.05	1.46	18.42	89.29	1.46	18.42	89.29
Bld	3.61	5.37	95.43	6.49	86.18	50.04	3.61	27.56	83.32	6.49	89.47	48.25
bb8	3.46	7.85	94.15	2.29	98.25	45.37	3.46	39.8	76.72	2.29	91.54	49.02
r2d2	2.20	12.36	92.26	1.66	96.34	46.70	2.20	57.93	67.06	1.66	89.02	50.69

Table 6. Results of LivDet 2021 competitors on test sets acquired with the Dermalog sensor. In particular, Derm CC is a test set acquired with the consensual technique, while Derm SS is a test set acquired with the ScreenSpoof technique. (*) indicates training on both GreenBit and Dermalog Livdet training sets. (**) indicates training on additional data with respect to LivDet training sets.

Alg.	Trained on DL and Tested on DL CC			Trained on GB and Tested on DL CC			Trained on DL and Tested on DL SS			Trained on GB and Tested on DL SS		
	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]
Col (**)	1.61	26.71	99.18	1.51	0.29	99.16	1.61	58.86	67.16	1.51	61.5	65.76
Dob (**)	1.07	0.16	99.37	1.27	0.20	99.31	1.07	31.34	82.41	1.27	26.59	84.92
con	5.27	0.28	93.46	36.44	0.16	83.35	5.27	73.94	57.27	36.44	45.50	58.07
JLW	0.68	30.65	98.16	5.51	25.08	83.81	0.68	95.12	45.41	5.51	99.92	43.00
m_s (*)	0.83	2.80	99.20	0.83	0.77	99.20	0.83	29.07	83.77	0.83	29.07	83.77
m_e (*)	0.24	0.77	99.87	0.24	0.04	99.87	0.24	28.66	84.26	0.24	28.66	84.26
PAD (*)	2.68	13.13	96.16	2.68	4.80	96.16	2.68	24.72	85.30	2.68	24.72	85.30
Bld	2.59	4.80	94.28	5.85	0.37	97.14	2.59	77.97	56.30	5.85	22.03	85.32
bb8	2.39	8.33	96.58	3.61	49.59	71.31	2.39	69.51	46.03	3.61	99.88	43.88
r2d2	1.27	4.27	98.03	0.73	68.29	62.42	1.27	82.11	45.85	0.73	100.00	45.12

To control the experiment more and to select data that were unknown during the training phase, we submitted images to the LivDet 2021 PADs acquired with two very different sensors compared to those used during the eighth edition of the competition, LivDet Orchantus 2019 and LivDet Biometrika 2013. The results of this analysis are shown in Table 7. In this case, some data are unavailable, so we have to restrict the analysis to six PADs. Apart from a few exceptions with *megvii*, all PADs show a significant drop in performance and are completely ineffective on the new data. This is evident, above all, from the results on Orchantus 2019, whose non-optical acquisition technology obtains very different images compared to the sensors of the training set. It is worth underlining that, based on the type of image, the error can be shifted entirely to fakes, as in the case of Biometrika 2013, or entirely to live, as in the case of Orcanthus 2019. From these results, we can hypothesize that training on different types of data improves the performance on unknown attacks such as ScreenSpoof at the expense of a greater intra-sensor APCER. This is evident, for example, from *LivDet_Dob_C2* in Table 5, which, in the intra-method test (i.e., trained on the GreenBit train and tested on the consensual GreenBit test (GB CC), is among the least accurate PADs with 85.88% accuracy, while in the cross-method test (i.e., trained on the GreenBit train and tested on the GreenBit ScreenSpoof (GB SS) test), it is the highest performing, exceeding 97% accuracy.

Table 7. Results of LivDet 2021 competitors on LivDet Biometrika 2013 and LivDet Orcanthus 2019 test sets. (*) indicates training on both GreenBit and Dermalog Livdet training sets.

Alg.	Trained on GB and Tested on BK 2013			Trained on DL and Tested on BK 2013			Trained on GB and Tested on OR 2019			Trained on DL and Tested on OR 2019		
	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]	BPCER [%]	APCER [%]	Liv. Acc. [%]
con	16.30	5.70	89.00	0.00	100.00	50.00	98.18	22.89	41.24	83.23	62.87	27.43
JLW	70.40	4.40	62.60	52.90	2.20	72.45	89.80	22.61	45.38	74.85	65.81	29.88
m_s	0.30	11.10	94.30	0.30	11.10	94.30	93.54	0.46	55.20	93.54	0.46	55.20
m_e	0.30	0.20	99.75	0.30	0.20	99.75	97.27	0.37	53.46	97.27	0.37	53.46
PAD (*)	0.00	92.40	53.80	0.00	92.40	53.80	53.64	39.34	53.85	53.64	39.34	53.85
Bld	18.17	34.00	73.65	16.10	97.70	43.10	99.80	0	52.45	81.21	10.02	56.06

However, the lack of controllability of the training phase is a limitation of this analysis. For this reason, in the next section, we analyzed the behavior of completely self-trained PADs where both the implementation and the training data were known.

4.3. Self-Trained Analysis

Starting from the findings reported in the previous section, we decided to carry out new experiments in which we carefully selected the training and test sets to evaluate different interoperable scenarios. We chose the same training sets of LivDet 2021 to compare the results directly and added the LivDet 2019 DigitalPersona dataset, keeping the same sensing technology of GreenBit and Dermalog scanners. We then merged these three datasets in different proportions while maintaining a regular training set size. Accordingly, we used half or one-third of the datasets when we combined two or three of them, respectively. This approach helps evaluate the impact of heterogeneous data on classification.

Thus, we designed three different PADs: two hand-crafted, based on a linear SVM classifier trained with two of the most adopted textural algorithms at the state of the art, namely BSIF and LBP [7,8]; and one based on the SpoofBuster algorithm [11]. We did not carry out any parameter optimization since our purpose was not to design the optimal PAD, but to enhance the improvement achievable when introducing samples coming from multiple scanners.

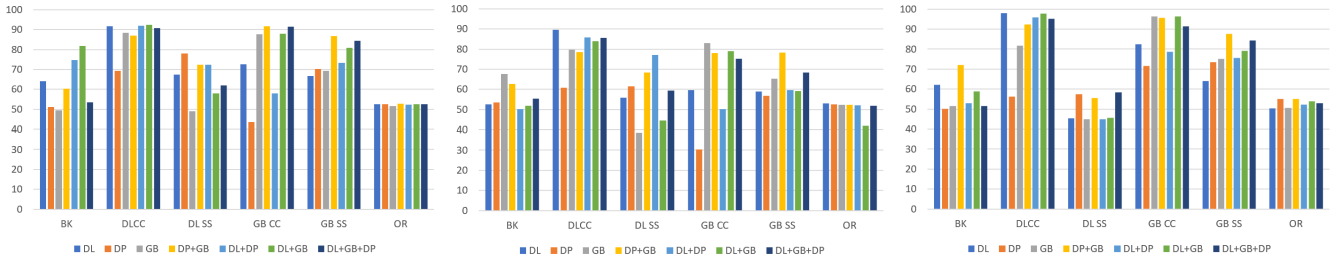
As a first analysis, we investigated the PADs' accuracies at a decision threshold of 0.5 score (range [0,1]) (Tables 8 and 9). Figure 3 illustrates the accuracy analysis visually to make it easy to read. From these results, it can be seen that, except for some tests related to LBP, the use of training on different types of data benefits the performance of the PADs. Again, this benefit is nil in the case of the Orcanthus 2019 test, in which all our self-trained PADs fail to classify the images. This is because this dataset is the only cross-sensing one. In fact, its PAs were acquired with thermal swipe technology, while all the training sets are related to optical technology, albeit with different sensors.

Table 8. Results in terms of accuracy, BPCER, and APCER with a threshold at 0.5 of BSIF, LBP, and SpoofBuster PADs with different single-sensor and multi-sensor training techniques for LivDet 2021 datasets.

	Scanner	DL CC			DL SS			GB CC			GB SS		
		BPCER	APCER	Accuracy	BPCER	APCER	Accuracy	BPCER	APCER	Accuracy	BPCER	APCER	Accuracy
BSIF	DL	4.93	11.02	91.75	4.93	55.85	67.29	3.56	47.11	72.68	3.56	58.25	66.61
	DP	42.63	20.89	69.22	42.63	4.76	78.03	19.56	87.36	43.46	19.56	38.25	70.24
	GB	8.49	14.31	88.34	8.49	86.50	48.96	4.49	18.98	87.61	4.49	52.64	69.25
	DL+GB	4.83	10.00	92.35	4.83	73.05	57.96	5.46	17.68	87.87	5.46	30.41	80.93
	DL+DP	5.66	9.96	92.00	5.66	45.77	72.46	2.39	75.16	57.92	2.39	46.95	73.30
	DP+GB	9.76	15.57	87.07	9.76	42.64	72.31	7.07	9.35	91.69	7.07	18.41	86.74
	DL+GB+DP	4.39	13.21	90.80	4.39	66.22	61.88	5.17	11.34	91.46	5.17	24.47	84.30
LBP	DL	5.80	14.15	89.65	5.80	76.26	55.76	37.02	43.17	59.62	37.02	44.59	58.85
	DP	84.93	1.10	60.80	84.93	0.04	61.37	49.17	86.87	30.27	49.17	38.13	56.85
	GB	24.83	16.59	79.67	24.83	92.03	38.51	8.59	23.90	83.06	8.59	56.54	65.25
	DL+GB	6.73	24.02	83.84	6.73	96.30	44.41	12.63	28.09	78.94	12.63	64.47	59.09
	DL+DP	11.27	16.50	85.88	11.27	32.64	77.07	15.95	78.29	50.04	15.95	60.61	59.69
	DP+GB	37.56	7.97	78.58	37.56	26.67	68.38	10.68	31.59	77.92	10.68	31.10	78.18
	DL+GB+DP	11.02	17.32	85.54	11.02	65.53	59.25	15.51	32.64	75.14	15.51	45.20	68.29
SpoofBuster	DL	1.22	2.60	98.03	1.22	99.23	45.32	1.71	30.57	82.55	2.83	63.37	64.15
	DP	90.49	4.76	56.27	90.49	2.64	57.42	48.24	11.67	71.71	51.12	5.89	73.55
	GB	1.36	32.15	81.84	1.37	99.80	44.94	2.34	4.71	96.36	2.14	43.82	75.12
	DL+GB	1.37	2.89	97.80	1.37	98.54	45.63	1.27	5.45	96.45	1.41	36.91	79.22
	DL+DP	1.02	6.54	95.96	1.02	99.88	45.05	8.93	31.50	78.76	12.54	34.39	75.54
	DP+GB	4.24	10.61	92.28	4.24	77.80	55.63	1.12	6.91	95.72	1.12	21.50	87.76
	DL+GB+DP	8.00	1.95	95.30	8.00	69.67	58.36	4.05	12.52	91.33	4.10	25.45	84.26

Table 9. Results in terms of accuracy, BPCER, and APCER with a threshold at 0.5 of BSIF, LBP, and SpoofBuster PADs with different single-sensor and multi-sensor training techniques for earlier LivDet editions datasets.

	Scanner	BK 2013			OR 2019		
		BPCER	APCER	Accuracy	BPCER	APCER	Accuracy
BSIF	DL	71.80	0.10	64.05	99.90	0.00	52.41
	DP	98.00	0.00	51.00	99.90	0.00	52.41
	GB	3.10	98.20	49.35	96.67	4.41	51.64
	DL+GB	33.00	3.20	81.90	99.70	0.00	52.50
	DL+DP	49.30	1.30	74.70	100.00	0.00	52.36
	DP+GB	7.30	72.30	60.20	99.29	0.00	52.69
	DL+GB+DP	0.50	92.70	53.40	99.90	0.00	52.41
LBP	DL	0.00	95.00	52.50	96.57	2.11	52.89
	DP	92.50	0.70	53.40	99.80	0.00	52.45
	GB	30.50	34.10	67.70	99.09	1.19	52.17
	DL+GB	0.00	96.30	51.85	89.49	29.50	41.92
	DL+DP	0.00	99.60	50.20	91.41	8.36	52.07
	DP+GB	43.20	31.50	62.65	100.00	0.00	52.36
	DL+GB+DP	0.30	89.00	55.35	95.25	5.24	51.88
SpoofBuster	DL	74.50	1.10	62.20	83.23	18.84	50.48
	DP	99.70	0.00	50.15	93.93	0.37	55.05
	GB	0.20	96.70	51.55	80.60	20.86	50.67
	DL+GB	80.60	1.50	58.95	94.14	2.48	53.85
	DL+DP	94.00	0.00	53.00	99.80	0.28	52.31
	DP+GB	42.20	13.50	72.15	93.84	0.46	55.05
	DL+GB+DP	96.70	0.10	51.60	98.38	0.18	53.03



(**a**) (**b**) (**c**)

Figure 3. Comparison of self-trained protocol accuracies for (**a**) BSIF, (**b**) LBP, and (**c**) SpoofBuster with different single-sensor and multi-sensor training techniques.

The next step has been to extend the investigation to all the liveness thresholds. The resulting analysis suggests that the advantages of employing multiple sensors depend on the final operating context. For a more precise reading, we have reported the errors related to some specific operational points in Tables 10 and 11, in particular, the BPCER value when the APCER is less than 5%, the APCER value when the BPCER is less than 5%, and the EER. This allows the simulation of application contexts in which first- and second-type errors weigh differently. The results indicate that, in an interoperable situation where avoiding presentation attacks is a primary concern, utilizing diverse data sources is ineffective since it increases the BPCER. In fact, in the intra-sensor and intra-method experiments, i.e.,

training and testing on data obtained from the same sensor and acquisition method, the use of additional data increases the error. Conversely, if the system is to be used in critical applications and the priority is to avoid rejecting genuine users, adding various sensors in the training set improves the spoof detection. This is especially evident when PADs are assaulted with never-seen-before attacks in the cross-method scenario, but it is also true in the intra-sensor and intra-method scenarios. For example, note the behavior of the PAD SpoofBuster when tested on consensual GreenBit data (GB CC); training on GreenBit data alone leads to an APCER at 5% and a BPCER equal to 9.88%; adding Dermalog data during the training phase drops the error to 6.41%. About the same improvement is obtained by adding DigitalPersona data.

ROC curves shown in Figures 4–9 give us more insights about these scenarios. They highlight that for the BSIF detector, training on GreenBit, Dermalog, and DigitalPersona is optimal at various operational points; for the SpoofBuster detector, the most effective combination is given by training on GreenBit and DigitalPersona.

It is also easy to see how the training can sometimes be ineffective; for instance, a BSIF-based PAD cannot distinguish the images coming from the Orcanthus sensor (Figure 5), and the LBP-based detector even reciprocates the classes, namely, it predicts a negative class as a positive class and vice versa, since the ROC is u-shaped. This case is frequent when DigitalPersona is used as the only dataset for training in combination with hand-crafted features. This is due to the great difference in the characteristics of the images acquired with the DigitalPersona sensor compared to those of GreenBit and Dermalog, as is evident from Figure 1. This behavior highlights how the same sensing technology does not always equate to a correspondence between the images' characteristics. In designing the PAD training phase, it is, therefore, necessary to know these characteristics to face a cross-sensor scenario. However, the minutiae-based approach of the FSB partly solves the difficulties of the textural algorithms, proving to be more efficient when trained on DigitalPersona, especially in the cross-scenarios. For this reason, we further investigate this method's behavior through its probability distributions (Figures 10–15).

Table 10. Results in terms of accuracy, BPCER, and APCER at different operational points of BSIF, LBP, and SpoofBuster PADs with different single-sensor and multi-sensor training techniques for LivDet 2021 datasets.

	Scanner	DL CC			DL SS			GB CC			GB SS		
		APCER (%)@ BPCER = 5%	BPCER (%)@ APCER = 5%	EER	APCER (%)@ BPCER = 5%	BPCER (%)@ APCER = 5%	EER	APCER (%)@ BPCER = 5%	BPCER (%)@ APCER = 5%	EER	APCER (%)@ BPCER = 5%	BPCER (%)@ APCER = 5%	EER
BSIF	DL	100.00	17.07	6.32	100.00	100.00	17.40	75.49	100.00	15.16	83.82	100.00	16.34
	DP	100.00	72.00	35.18	100.00	62.39	20.78	100.00	100.00	53.42	100.00	100.00	28.83
	GB	100.00	100.00	11.00	100.00	100.00	50.42	57.85	14.05	6.57	80.28	100.00	16.28
	DL+GB	100.00	18.63	6.48	100.00	100.00	27.79	60.89	17.32	8.03	70.24	100.00	11.93
	DL+DP	65.24	23.27	6.92	86.71	100.00	19.28	88.86	100.00	25.88	61.46	100.00	14.19
	DP+GB	100.00	48.39	11.41	100.00	100.00	17.85	72.76	17.66	7.87	74.27	24.63	9.96
	DL+GB+DP	62.97	21.95	6.67	96.14	100.00	17.81	60.20	15.37	6.79	68.01	29.90	10.84
LBP	DL	61.67	23.22	9.04	94.27	100.00	32.00	100.00	100.00	38.78	100.00	100.00	39.83
	DP	100.00	85.27	42.80	100.00	80.68	40.61	100.00	100.00	67.33	100.00	100.00	44.29
	GB	100.00	60.20	21.73	100.00	100.00	58.45	100.00	100.00	13.73	100.00	100.00	23.01
	DL+GB	100.00	40.05	12.27	100.00	100.00	48.85	100.00	100.00	16.63	100.00	100.00	29.11
	DL+DP	81.22	37.66	12.63	86.38	100.00	18.66	100.00	100.00	35.97	100.00	100.00	33.59
	DP+GB	100.00	61.85	25.95	100.00	100.00	31.79	76.10	100.00	17.83	70.69	100.00	18.11
	DL+GB+DP	100.00	44.73	12.98	100.00	100.00	30.11	100.00	100.00	20.07	100.00	100.00	24.79
SpoofBuster	DL	3.98	2.20	1.78	99.51	100.00	24.22	41.26	100.00	9.43	86.99	77.80	21.46
	DP	100.00	96.34	51.09	100.00	93.27	35.58	97.20	100.00	29.57	100.00	76.83	24.50
	GB	36.79	33.95	8.75	99.92	100.00	58.74	9.88	100.00	3.55	62.64	60.49	12.39
	DL+GB	5.69	3.56	1.95	99.51	100.00	26.35	6.14	100.00	3.40	43.09	57.12	9.81
	DL+DP	6.54	5.80	2.80	99.88	100.00	34.83	84.51	100.00	16.87	87.89	73.12	21.10
	DP+GB	33.25	22.88	6.71	95.20	56.29	22.66	7.20	100.00	3.57	28.01	38.24	6.21
	DL+GB+DP	24.27	12.49	4.38	99.80	53.32	22.25	43.70	100.00	7.51	62.64	35.80	9.09

Table 11. Results in terms of accuracy, BPCER, and APCER at different operational points of BSIF, LBP, and SpoofBuster PADs with different single-sensor and multi-sensor training techniques for earlier LivDet editions datasets.

	Scanner	BK 2013			OR 2019		
		APCER (%)@ BPCER = 5%	BPCER (%)@ APCER = 5%	EER	APCER (%)@ BPCER = 5%	BPCER (%)@ APCER = 5%	EER
BSIF	DL	100.00	62.50	31.50	100.00	99.49	49.84
	DP	100.00	93.00	46.50	100.00	99.60	49.80
	GB	99.30	100.00	50.45	100.00	98.59	51.27
	DL+GB	100.00	45.40	18.40	100.00	98.89	49.49
	DL+DP	100.00	53.10	20.70	100.00	100.00	50.00
	DP+GB	100.00	100.00	28.60	100.00	97.98	48.99
	DL+GB+DP	89.70	100.00	38.65	100.00	96.36	48.18
LBP	DL	90.80	100.00	45.55	100.00	97.68	51.83
	DP	100.00	91.90	45.90	100.00	99.80	49.90
	GB	100.00	100.00	32.95	100.00	99.49	54.36
	DL+GB	100.00	100.00	46.30	100.00	100.00	67.99
	DL+DP	100.00	100.00	49.50	100.00	97.78	54.94
	DP+GB	100.00	100.00	35.25	100.00	97.07	49.09
	DL+GB+DP	85.20	100.00	42.45	100.00	98.99	59.95
SpoofBuster	DL	95.20	75.50	31.15	100.00	97.58	50.18
	DP	100.00	87.20	32.25	100.00	91.41	28.26
	GB	87.60	100.00	21.65	98.90	99.70	51.09
	DL+GB	100.00	84.50	39.55	100.00	97.17	42.72
	DL+DP	100.00	79.60	37.45	100.00	99.60	49.01
	DP+GB	92.70	74.70	28.05	94.03	88.69	32.62
	DL+GB+DP	100.00	87.30	46.60	100.00	94.65	47.65

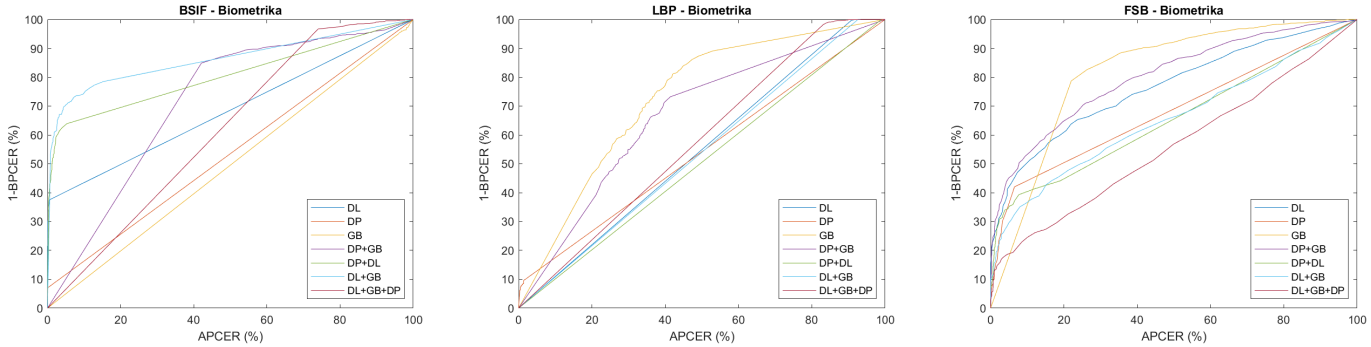


Figure 4. Comparison between ROCs of our self-trained PAD systems in an intra- (dashed) and cross- (solid) sensor scenario on the Biometrika dataset from LivDet 2013.

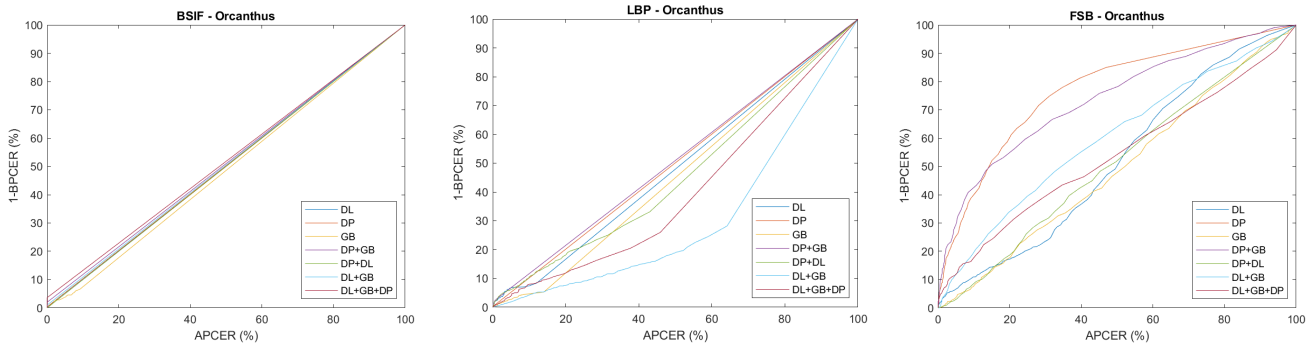


Figure 5. Comparison between ROCs of our self-trained PAD systems in an intra- (dashed) and cross- (solid) sensor scenario on the Orcanthus dataset from LivDet 2019.

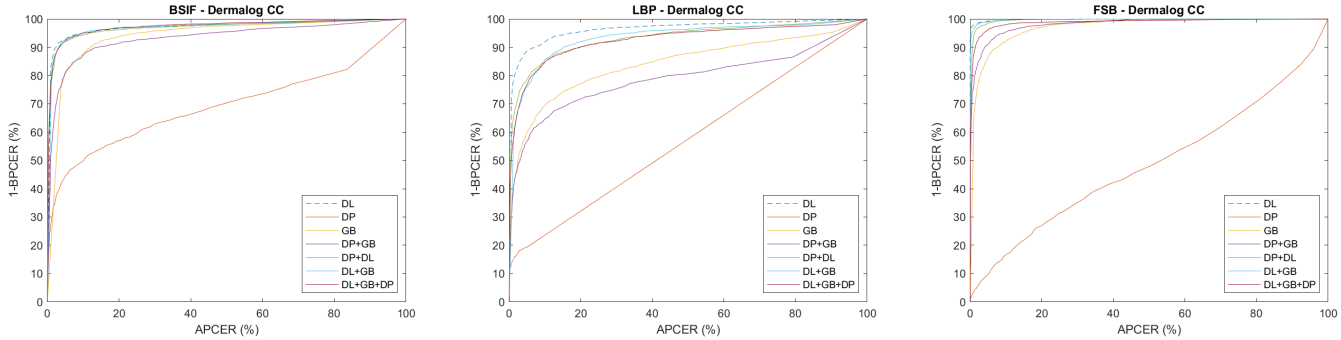


Figure 6. Comparison between ROCs of our self-trained PAD systems in an intra- (dashed) and cross- (solid) sensor scenario on the Dermalog Consensual dataset from LivDet 2021.

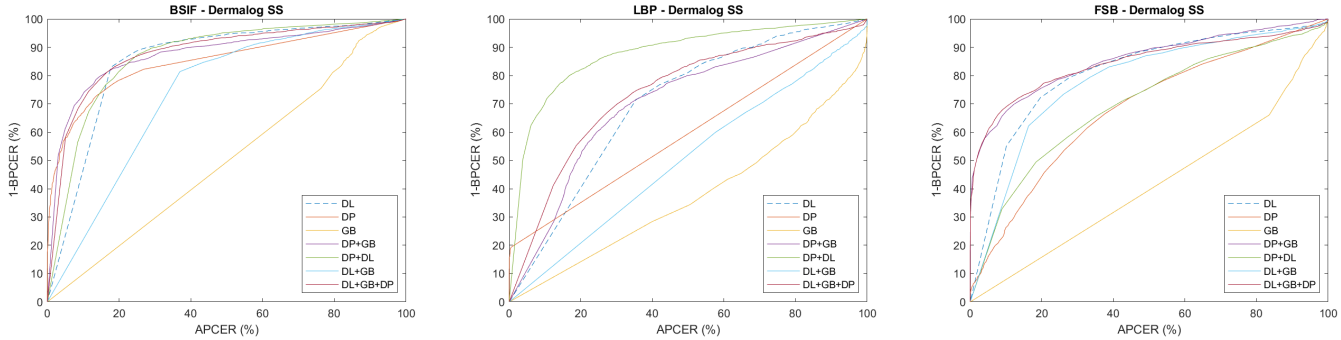


Figure 7. Comparison between ROCs of our self-trained PAD systems in an intra- (dashed) and cross- (solid) sensor scenario on the Dermalog ScreenSpoof dataset from LivDet 2021.

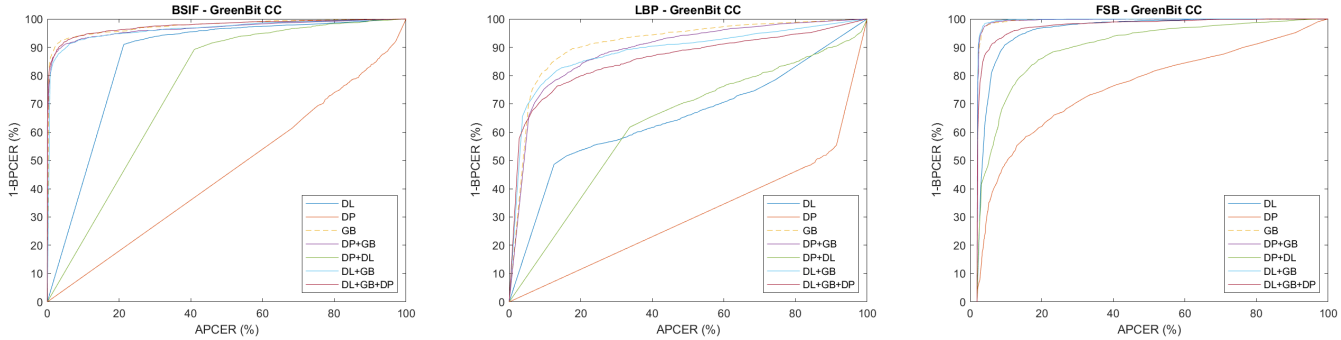


Figure 8. Comparison between ROCs of our self-trained PAD systems in an intra- (dashed) and cross- (solid) sensor scenario on the GreenBit consensual dataset from LivDet 2021.

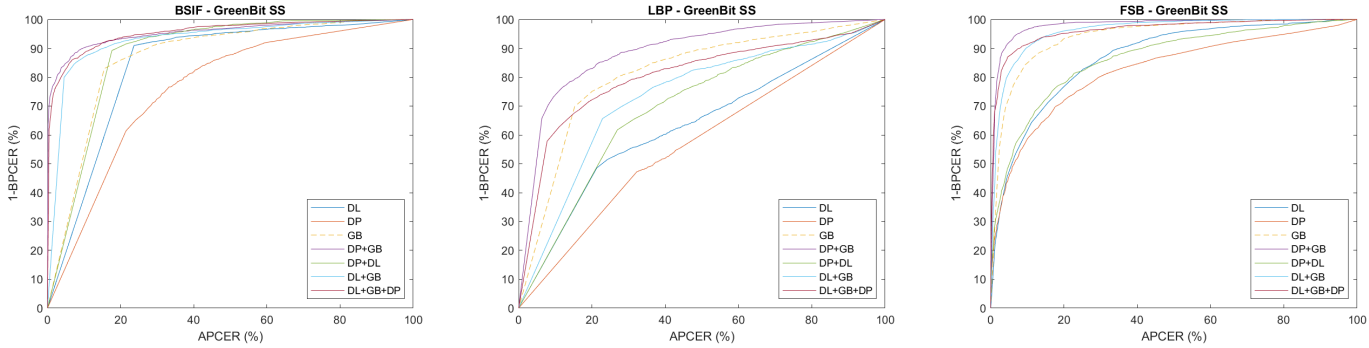
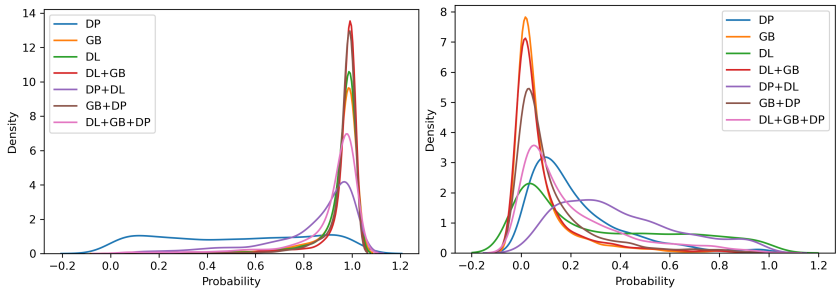
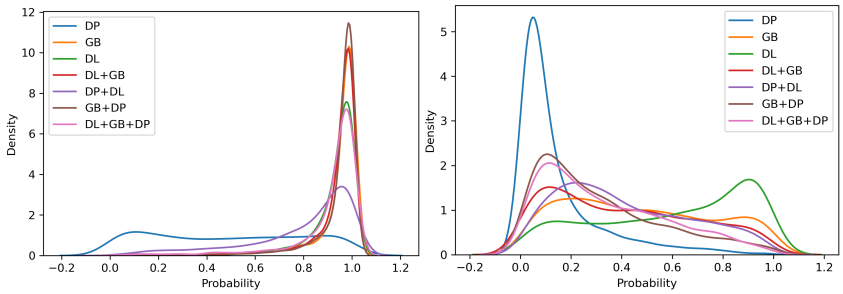


Figure 9. Comparison between ROCs of our self-trained PAD systems in an intra- (dashed) and cross- (solid) sensor scenario on the GreenBit ScreenSpoof dataset from LivDet 2021.



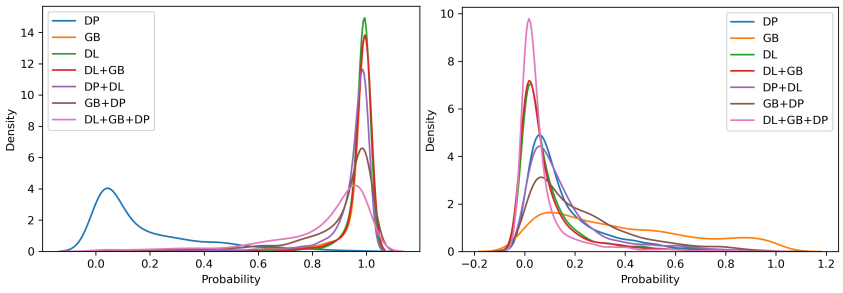
(**a**) (**b**)

Figure 10. Probability distributions of (**a**) real and (**b**) fake fingerprint scores of the GreenBit consensual dataset obtained by the PAD Spoofbuster.



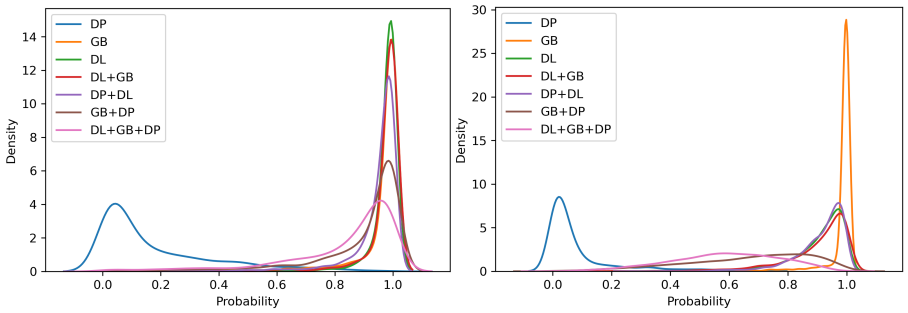
(**a**) (**b**)

Figure 11. Probability distributions of (**a**) real and (**b**) fake fingerprint scores of the GreenBit ScreenSpoof dataset obtained by the PAD Spoofbuster.



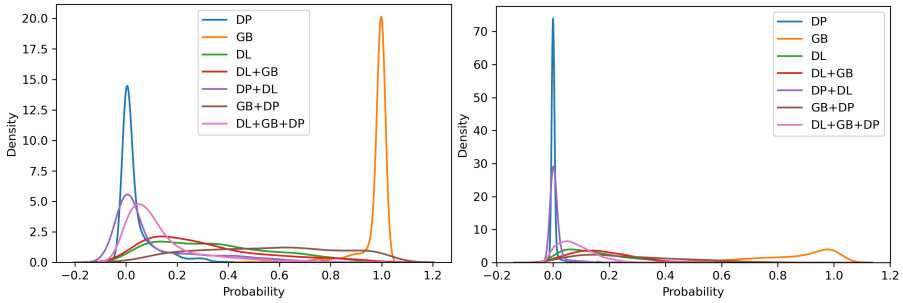
(**a**) (**b**)

Figure 12. Probability distributions of (**a**) real and (**b**) fake fingerprint scores of the Dermalog consensual dataset obtained by the PAD Spoofbuster.



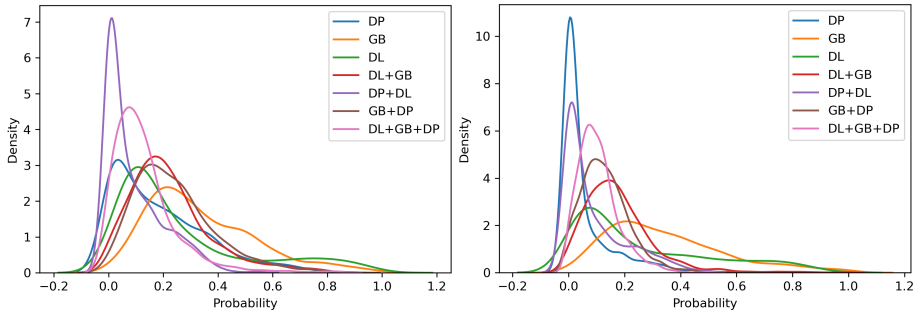
(**a**) (**b**)

Figure 13. Probability distributions of (**a**) real and (**b**) fake fingerprint scores of the Dermalog Screen-Spoof dataset obtained by the PAD Spoofbuster.



(**a**) (**b**)

Figure 14. Probability distributions of (**a**) real and (**b**) fake fingerprint scores of the Biometrika 2013 dataset obtained by the PAD Spoofbuster.



(**a**) (**b**)

Figure 15. Probability distributions of (**a**) real and (**b**) fake fingerprint scores of the Orchantus 2019 dataset obtained by the PAD Spoofbuster.

Optimal distributions are represented by unimodal distributions with low variability, averaging equal to zero for lives and equal to one for fakes. Besides confirming the observation previously presented, this representation allows us to appreciate the impact of training with multiple data on the classification. The use of multiple data in the training phase allows us, in fact, to obtain mono-modal distributions using datasets that individually lead to multi-modal distributions or distributions with averages far from the optimal ones.

This is evident from the GreenBit ScreenSpoof data distributions (Figure 11), in which the model trained on GreenBit and DigitalPersona positively exploits the GreenBit contribution on lives and the DigitalPersona contribution on fakes, allowing it to obtain a minimum overlap with respect to the distributions of the single models (brown curve). Furthermore, this analysis allows us to explain why fusion does not influence cross-sensor scenarios with different sensing technology, as in the case of the Orchantus experiments. The samples

acquired with the Orchantus sensor, in fact, always correspond to a low output score for all the tested models. The models cannot represent the two classes and all the samples are classified as fake. The fusion, therefore, fails to bring benefits because the information at the base is missing; there are no representative samples in the training phase. This shows that the benefit of training on different types of data must be adequately designed to cover all the types of data that the system must be able to recognize; if data are not available on a specific sensor to be integrated into an AFIS, it is necessary to find data as close as possible in terms of image characteristics to those of the target sensor. This is also demonstrated by the results on Dermalog ScreenSpoof (DL SS). Training on the DigitalPersona and GreenBit combination obtains an EER almost equal to or less than training on Dermalog data. The GreenBit data are very similar in terms of image characteristics to those of Dermalog. Using DigitalPersona data increases the ability to generalize and better recognize cross-method data.

5. Discussion and Conclusions

In this paper, we analyzed the impact on cross-sensor use of introducing variety in the type of training data, i.e., when the test data has been acquired with a different sensor than the training data; and on cross-method use, i.e., when the test data have been acquired with different techniques. This allowed us to provide some insights to PAD designers based on cost/efficiency trade-offs and the specific application context of the system. We have reported the main findings of our investigation in Table 12, from which we can derive the following guidelines:

- For intra-method and intra-sensor experiments, training on the target sensor is preferable; however, training on multiple sensors does not significantly worsen the results;
- For cross-method experiments, training on different types of images allows obtaining better results for operational points relative to low APCERs. In general, using numerous data for ScreenSpoof tests at the EER is comparable to or better than the single best training;
- For the cross-sensor experiments, it is not possible to detect a benefit related to the use of training on multiple sensors. However, even single-sensor training does not result in effective PADs, showing that the interoperability problem is still open, and it is not possible to solve it without references from the target sensor. In particular, the need to use in training =the same sensing technology for data acquisition with respect to that expected during system operation was highlighted.

To sum up, if the usage of the AFIS is entirely controlled, there is no expectation of sensor change over time and there is a low probability of falsification using unknown techniques. It is preferable to train with specific target data. If the aim is to make the system more robust to never-before-seen attacks, it is preferable to train it on different data sources. However, an analysis of the image characteristics is necessary to select the best training data.

This work also highlights the need for an in-depth study of the characteristics of both live and fake fingerprint images to obtain a representation capable of including all intra-class variations due to different factors such as sensing technology, sensor size, materials, techniques for making a fake (for PAs), or skin conditions (for lives).

Table 12. Pros and cons of training on data acquired with multiple sensors for the investigated scenarios.

Scenario		Property	Multi-Sensor Training Pros	Multi-Sensor Training Cons
Intra-sensor and intra-method		• Same sensor in training and test sets; • Same method of spoof fabrication.	-	• Increase the BPCER; • Effectiveness depends on the final operating context.
Intra-sensor and cross-method		• Same sensor in training and test sets; • Different methods of spoof fabrication.	• More robust to PA committed with unknown methods; • Better results for operational points relative to low APCERs.	-
Cross-sensor and intra-method	Intra-sensing	• Different sensors in training and test sets, but with the same sensing technology (e.g., all optical); • Same method of spoof fabrication.	• Low training costs; • Beneficial, especially on APCER.	• Improvement depends on the quality of the PAD.
	Cross-sensing	• Different sensors in training and test sets, with different sensing technologies (e.g., optical and capacitive); • Same method of spoof fabrication.	-	• Ineffective: images could be very different, leading to a low accuracy.
Cross-sensor and cross-method	Intra-sensing	• Different sensors in training and test sets, but with the same sensing technology (e.g., all optical); • Different methods of spoof fabrication.	• Low training costs; • General improvement with multi-sensor training; the BPCER at stringent operational thresholds improves considerably; • Better accuracy then single sensor training.	-
	Cross-sensing	• Different sensors in training and test sets, with different sensing technologies (e.g., optical and capacitive); • Different methods of spoof fabrication.	Not evaluated.	

Author Contributions: Conceptualization, G.L.M.; methodology, M.M., G.O., and R.C.; software, M.M., G.O., and R.C.; validation, M.M., G.O., and R.C.; formal analysis, M.M., G.O., and R.C.; investigation, M.M., G.O., and R.C.; resources, M.M., G.O., and R.C.; data curation, R.C.; writing—original draft preparation, M.M., G.O., and R.C.; writing—review and editing, M.M., G.O., R.C., and G.L.M.; supervision, G.L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data supporting reported results can be found at https://livdet.diee.unica.it/ (accessed on 25 September 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PAD	Presentation Attack Detection
FPAD	Fingerprint Presentation Attack Detection
SIFS	Scale Invariant Feature Transformation
BSIF	Binarized Statistical Image Features
LBP	Local Binary Pattern
LPQ	Local Phase Quantization

SVM	Support Vector Machine
CNN	Convolutional Neural Networks
APCER	Attack Presentation Classification Error Rate
BPCER	Bona fide Presentation Classification Error Rate
ROC	Receiver Operating Characteristic
GB	GreenBit dataset—LivDet 2021 train
GB CC	GreenBit consensual dataset—LivDet 2021 test
GB SS	GreenBit ScreenSpoof dataset—LivDet 2021 test
DL	Dermalog dataset—LivDet 2021 train
DL CC	Dermalog consensual dataset—LivDet 2021 test
DL SS	Dermalog ScreenSpoof dataset—LivDet 2021 test
BK	Biometrika dataset—LivDet 2013 test
OR	Orcanthus dataset—LivDet 2019 test
DP	DigitalPersona dataset—LivDet 2019 train

References

1. Chugh, T.; Jain, A.K. Fingerprint spoof detector generalization. *IEEE Trans. Inf. Forensics Secur.* **2020**, *16*, 42–55. [CrossRef]
2. Casula, R.; Orrù, G.; Angioni, D.; Feng, X.; Marcialis, G.L.; Roli, F. Are spoofs from latent fingerprints a real threat for the best state-of-art liveness detectors? In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3412–3418.
3. Sharma, D.; Selwal, A. FinPAD: State-of-the-art of fingerprint presentation attack detection mechanisms, taxonomy and future perspectives. *Pattern Recognit. Lett.* **2021**, *152*, 225–252. [CrossRef]
4. Willis, D.; Lee, M. Six Biometric Devices Point the Finger at Security. *Netw. Comput.* **1998**, *9*, 84–96.
5. Kallo, P.; Kiss, I.; Podmaniczky, A.; Losi, J. Detector for Recognizing the Living Character of a Finger in a Fingerprint Recognizing Apparatus. U.S. Patent 6,175,641, 16 January 2001.
6. Schuckers, S.A.C. Spoofing and anti-spoofing measures. *Inf. Secur. Tech. Rep.* **2002**, *7*, 56–62. [CrossRef]
7. Ghiani, L.; Hadid, A.; Marcialis, G.L.; Roli, F. Fingerprint liveness detection using binarized statistical image features. In Proceedings of the 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 29 September–2 October 2013; pp. 1–6.
8. Gragnaniello, D.; Poggi, G.; Sansone, C.; Verdoliva, L. Fingerprint liveness detection based on weber local image descriptor. In Proceedings of the 2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications, Napoli, Italy, 9 September 2013; pp. 46–50.
9. Rattani, A.; Scheirer, W.J.; Ross, A. Open Set Fingerprint Spoof Detection Across Novel Fabrication Materials. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 2447–2460. [CrossRef]
10. Grosz, S.A.; Chugh, T.; Jain, A.K. Fingerprint presentation attack detection: A sensor and material agnostic approach. In Proceedings of the 2020 IEEE International Joint Conference on Biometrics (IJCB), Houston, TX, USA, 28 September–1 October 2020; pp. 1–10.
11. Chugh, T.; Cao, K.; Jain, A.K. Fingerprint spoof buster: Use of minutiae-centered patches. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2190–2202. [CrossRef]
12. Lugini, L.; Marasco, E.; Cukic, B.; Gashi, I. Interoperability in fingerprint recognition: A large-scale empirical study. In Proceedings of the 2013 43rd Annual IEEE/IFIP Conference on Dependable Systems and Networks Workshop (DSN-W), Budapest, Hungary, 24–27 June 2013; pp. 1–6.
13. Alshehri, H.; Hussain, M.; Aboalsamh, H.A.; Emad-Ul-Haq, Q.; AlZuair, M.; Azmi, A.M. Alignment-free cross-sensor fingerprint matching based on the co-occurrence of ridge orientations and Gabor-HoG descriptor. *IEEE Access* **2019**, *7*, 86436–86452. [CrossRef]
14. Marasco, E.; Sansone, C. On the Robustness of Fingerprint Liveness Detection Algorithms against New Materials used for Spoofing. *Biosignals* **2011**, *8*, 553–555.
15. Gajawada, R.; Popli, A.; Chugh, T.; Namboodiri, A.; Jain, A.K. Universal material translator: Towards spoof fingerprint generalization. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019; pp. 1–8.
16. Tuveri, P.; Ghiani, L.; Zurutuza, M.; Mura, V.; Marcialis, G.L. Interoperability among capture devices for fingerprint presentation attacks detection. In *Handbook of Biometric Anti-Spoofing*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 71–108.
17. ISO/IEC 30107-3:2017(en). Information Technology-Biometric Presentation Attack Detection-Part 3: Testing and Reporting. 2017. Available online: https://www.iso.org/obp/ui/#iso:std:iso-iec:30107:-3:ed-1:v1:en (accessed on 19 September 2022).
18. Micheletto, M.; Orrù, G.; Casula, R.; Yambay, D.; Marcialis, G.L.; Schuckers, S.C. Review of the Fingerprint Liveness Detection (LivDet) competition series: From 2009 to 2021. *arXiv* **2022**, arXiv:2202.07259.