LivDet2025: Toward Robust and Generalizable Fingerprint Presentation Attack Detection

Giulia Orrù[1], Marco Micheletto[1], Roberto Casula[1], Simone Zedda[1], Daniele Fenu[1], Lambert Igene[2], Jannis Priesnitz [3], Christoph Busch [3,4], Christian Rathgeb[3], Stephanie C. Schuckers [5], Gian Luca Marcialis[1]
[1] University of Cagliari, Italy, [2] Clarkson University, USA,
[3] Hochschule Darmstadt, Germany, [4] Norwegian University of Science and Technology, Norway
[5] The University of North Carolina at Charlotte, USA
Corresponding email: giulia.orru@unica.it

Abstract

The Fingerprint Liveness Detection Competition (LivDet) is a recurring benchmark series that evaluates the effectiveness of software-based Presentation Attack Detection (PAD) algorithms in fingerprint recognition. LivDet2025 presents three challenges: (1) "Liveness Detection in Action", requiring the integration of PAD with user-specific recognition; (2) "Fingerprint Representation", evaluating the compactness and discriminability of feature vectors; and (3) "Adversarial Robustness", assessing the resilience of PADs to adversarially-crafted presentation attack instruments. This edition marks a significant milestone with the inclusion of contactless fingerprint data, promoting interoperability and robustness across acquisition technologies. Furthermore, no training data was provided; participants must select and declare external datasets for model development. The competition was open to academic and industrial research groups, with all submitted algorithms evaluated on common datasets and under standardized protocols. LivDet2025 aims to provide a comprehensive assessment of PAD performance under realistic, multi-sensor, and multi-attack scenarios. Results reveal important trade-offs between PAD accuracy, usability, and computational efficiency. For instance, some systems achieved high presentation attack rejection at the cost of extremely high false rejection rates, while others optimised speed and generalizability but exhibited limited attack resilience.

1. Introduction

Fingerprint-based authentication remains a cornerstone of biometric security systems, for its uniqueness, permanence, and operational efficiency [17]. However, its large-scale deployment has also exposed inherent vulnerabilities—most notably, presentation attacks (PAs), in which adversaries employ synthetic artifacts (e.g., silicone, gelatin, printed overlays) to impersonate legitimate users and deceive acquisition devices [12]. Software-based Presentation Attack Detection (PAD) methods have been extensively developed to address this threat, leveraging image-derived features and machine learning algorithms to differentiate bona fide from presentation attack (i.e. spoof) samples [3]. Despite recent progress, several challenges continue to hinder the reliable deployment of PAD systems in practical scenarios [23]. A first, well-documented issue is their limited generalization capability: performance often degrades significantly when algorithms are tested against presentation attack instrument species or sensors not encountered during training [14]. In parallel, the growing reliance on machine learning has introduced vulnerabilities to adversarial manipulation, subtle, targeted perturbations, either in digital form or physically realizable, that can induce erroneous predictions from otherwise accurate detectors [6]. A third obstacle lies in the computational and memory footprint of many PAD solutions, which makes their deployment infeasible on resource-constrained platforms such as mobile or embedded devices [16]. In addition to these algorithmic limitations, the contact-based nature of traditional fingerprint acquisition presents further usability challenges. Direct interaction between the finger and sensor surface often results in degraded image quality due to moisture, dirt, or latent fingerprints from prior users. Moreover, in shared-use scenarios, hygiene and pathogen transmission concerns have been shown to impact user acceptance negatively [21]. This has motivated research into contactless fingerprint acquisition, where no physical touch is required. Systems based on smartphone or optical cameras offer hygienic and ergonomic advantages but also introduce new challenges, including pose variability, illumination inconsistency, and

 DOI: 10.1109/IJCB65343.2025.11411004

reduced ridge contrast [8]. Additionally, contactless systems face novel attack vectors such as ScreenSpoof attacks [5], which differ substantially from threats in traditional PAD scenarios [15]. Given these evolving conditions, it becomes essential to evaluate PAD solutions not only in controlled laboratory settings but across a range of acquisition modalities and threat models. Since 2009, the Fingerprint Liveness Detection Competition (LivDet[1]) has served as the principal benchmark for software-based PAD algorithms, focusing on contact-based fingerprints due to their operational prevalence [19]. Previous editions have emphasized two main challenges: (1) the integration of PAD within full recognition pipelines and (2) the extraction of compact, discriminative feature representations. LivDet2025 builds on this foundation by introducing two targeted extensions. First, a dedicated contactless track has been added to evaluate PAD performance on contactless fingerprint data. Second, motivated by growing concerns over adversarial robustness, a new adversarial challenge has been added within the contact-based track to evaluate algorithmic resilience under targeted digital perturbations. In addition, and for the first time in the competition's history, LivDet2025 does not provide participants with predefined training data. Each team is instead required to develop models using publicly available datasets of their choice and to disclose the sources employed. The objective is to shift the focus toward genuine generalization, discouraging solutions narrowly tuned to the benchmark. By expanding its scope and adopting a more realistic evaluation protocol, LivDet2025 encourages the design of PAD systems that can operate reliably across diverse acquisition conditions and evolving attack landscapes.

2. LivDet2025

LivDet2025 includes three challenges. Challenge 1 and Challenge 2 are inherited from previous editions and are evaluated separately on contact-based and contactless fingerprints. Participants may choose to compete in one or both. Challenge 3 is newly introduced, and applies to all submitted systems, and focuses on adversarial robustness using contact-based data.

- Challenge 1 - *Liveness Detection in Action* [7]: Participants are required to submit a complete system that outputs both a liveness score (i.e., the probability that the sample is bona fide) and an integrated score that combines liveness with the likelihood of the sample belonging to a specific user. The use of user-specific models [20] is optional.
- Challenge 2 - *Fingerprint representation*: To promote efficient and scalable PAD systems, this challenge evaluates the ability to produce feature vectors that are both compact and discriminative. In addition to the liveness score, each submission must return a fixed-length feature representation of the input image, with a maximum dimensionality of 512.
- Challenge 3 - *Adversarial Robustness*: This newly introduced track targets the resilience of PAD systems against adversarial attacks, i.e. intentionally crafted digital perturbations aimed at inducing misclassification. Solutions are evaluated based on their ability to maintain accuracy under these conditions.

2.1. Datasets and participants

Table 3 provides detailed information about each participant, including the names of the algorithms submitted, the type of method used, and the specific data sets used for training. In fact, in this edition, unlike previous ones, competitors were required to select state-of-the-art datasets for model training and explicitly declare the data used, as no training data were provided.

The LivDet2025 evaluation set comprises both contact-based and contactless fingerprint images, acquired with a total of six sensors across the two modalities. The dataset includes bona fide fingerprints and presentation attacks fabricated using a variety of materials and methods, aiming to capture realistic deployment scenarios. Sensor specifications and the distribution of samples across materials and devices are reported in Table 1 and Table 2, respectively. Figure 1 shows representative PA samples highlighting the variability introduced by acquisition conditions and fabrication techniques. The dataset is structured into two distinct subsets according to the acquisition modality.

Contact-based fingerprints were acquired using two optical scanners: *GreenBit DactyScan84C* and *Dermalog LF10*. For each device, the dataset comprises bona fide samples as well as presentation attack samples created through both consensual and semi-consensual interactions. Alongside presentation attack instruments (PAIs) fabricated with traditional consensual techniques, a significant portion of the attacks was produced using the *ScreenSpoof* method [5], a semi-consensual technique that reconstructs fingerprint molds from latent traces photographed on smartphone screens, which are then used to fabricate PAIs.

Contactless acquisitions were performed using four consumer-grade mobile devices: Huawei P20 Pro (collected at the University of Cagliari) and iPhone 7, iPhone X, and Samsung Galaxy S7 (collected at Clarkson University [22]). The datasets comprise bona fide samples and presentation attacks fabricated using traditional consensual methods, with materials including latex, ecoflex, playdoh, wood glue, and printed photopaper. The acquisitions reflect a wide range of real-world conditions, including uncontrolled lighting, pose variability, background clutter, and different acquisition distances.

[1] https://sites.unica.it/livdet/

Table 1. Device characteristics for LivDet2025 datasets. For smartphones, the reported image resolution corresponds to the native sensor output. Actual fingerprint images may be cropped or resized.

Device	Acronym	Model / Camera	Resolution [dpi]	Sensor Resolution [px]	Type
Green Bit	GB	DactyScan84C, 500 dpi	500	500×500	Optical
Dermalog	DL	LF10, 500 dpi	500	500×500	Optical
Huawei P20 Pro	HW	Rear Camera, 40 MP	72–96	7152×5368	Smartphone
iPhone 7	I7	Rear Camera, 12 MP	72–96	4608×2592	Smartphone
iPhone X	IX	Rear Dual Camera, 12 MP	72–96	4608×2592	Smartphone
Samsung Galaxy S7	S7	Rear Camera, 12 MP	72–96	4032×3024	Smartphone

Table 2. Number of samples per material and sensor in LivDet2025. Materials marked with * are *ScreenSpoof* samples.

Sensor → / **Material ↓**	**Contact**		**Contactless**			
	GreenBit	Dermalog	Huawei	iPhone 7	iPhone X	Samsung Galaxy S7
Bonafide	1750	1750	720	592	477	788
Elmers	660	656	288	·	·	·
Latex	680	680	252	1120	1120	·
Rprotec	641	630	·	·	·	·
GLSpro*	650	648	·	·	·	·
GSP400n*	672	668	·	·	·	·
Latex*	670	667	·	·	·	·
Prolastix*	670	668	·	·	·	·
Ecoflex	·	·	·	1120	1112	·
Photopaper	·	·	·	1120	1120	·
Playdoh	·	·	288	384	384	367
Woodglue	·	·	·	1120	1120	·

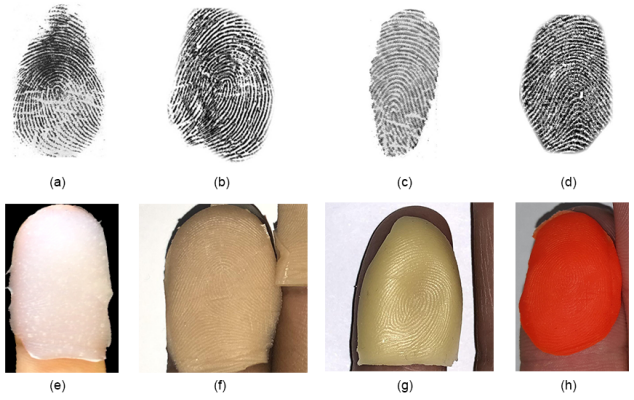

Figure 1. Examples of presentation attacks (PAs) used in LivDet2025. Images illustrate visual and material variability across sensors and acquisition conditions. From left to right: (a)–(b) optical contact-based Consensual samples (GreenBit, Dermalog); (c)–(d) optical contact-based ScreenSpoof samples (GreenBit, Dermalog); (e) contactless sample from Huawei; (f)–(h) contactless PAs from iPhone and Samsung devices.

2.2. Algorithms submission

Algorithms for Challenge 1 were expected as console programs with the following parameters:
[nameOfAlgorithm] [templateimagesfile] [probeimagesfile] [livenessoutputfile] [IMSoutputfile].
The file *templateimagesfile* contains a list of absolute paths to every template image stored in the system, while the file *probeimagesfile* contains a list of absolute paths to each probe image that the algorithm will test. The algorithm outputs are saved to the paths specified by the last two parameters. The file *livenessoutputfile* contains the degree of "liveness" for each processed image, normalized between 0 and 100, where 100 indicates the highest degree of liveness, and 0 denotes a presentation attack sample. Fingerprint images with scores in the range [0,50) are classified as "presentation attack", while those with scores [50,100] are classified as "bona fide". The file *IMSoutputfile* lists, for each probe image, the normalized probability of a fingerprint belonging to the declared identity and being authentic. Scores [0,50) classify the probe as "presentation attack" or the probe-template comparison as non-mated comparison, while scores [50,100] classify the comparison as bona-fide and mated. The evaluation threshold is set to 50. If the algorithm is unable to process an image, the corresponding value in both outputs is set to -1000.

The submission process for Challenge 2 in LivDet2025 is the same as in LivDet 2023. In addition to the parameters *nameOfAlgorithm*, *probeimagesfile*, and *livenessoutputfile*, Challenge 2 applications require an additional parameter called *embeddingsfile*, representing the file of feature vectors for each processed image.

Table 3. Participants' names and submitted algorithms, alongside information about their training approach.

Participant	Algorithm	Type	Training data	Description
UNIFESP	UNIFESP	Hybrid	LivDet 2009, 2011, 2013, 2015, 2017 [19]	Adaptation of Contreras et al.'s [9, 10] multi-filter approach using frozen pre-trained deep networks as feature extractors.
Dermalog	Derm_C	Deep-learning	LivDet 2021 [4], Internal Dermalog LF10 dataset	Extracts minutiae, generates cutouts, encodes them with MobileNetV2, and fuses results with an RNN.
	Derm_CL	Deep-learning	COLFISPOOF 2023 [15], LivDet contactless fingerprints [22], internal contactless data	Uses two MobileNetV2 models: one for fingertip shape verification and one for PA detection via classification and segmentation.
JIIOV Technology	jiiov	Deep-learning	LivDet 2009, 2011, 2013, 2015, 2021, 2023 [19], synthetic data via generative model, internal contact-based data	CNN-based keypoint extraction for fingerprint matching and transformer-based PAD model trained with real, synthetic, and internal data.

2.3. Performance Evaluation

In both challenges, the performance of the PADs will be evaluated using metrics from the international standard ISO/IEC 3017-3 [2, 1]:

- PAD Accuracy: percentages of fingerprint images correctly classified by the PAD;
- BPCER (Bona fide Presentation Classification Error Rate): Rate of misclassified bona fide images;
- APCER (Attack Presentation Classification Error Rate): Rate of misclassified presentation attack images;
- FTX (Failure to Extract): indicates failure in extracting biometric features.

In Challenge 1, to evaluate the performance of the integrated system, we employed the following metrics:

- FNMR (False Non-Match Rate): Rate of mated comparisons that result in rejection;
- FMR (False Match Rate): Rate of non-mated comparisons that result in acceptance;
- IAPAR (Impostor Attack Presentation Accept Rate): rate of presentation attacks that result in acceptance;
- Integrated Matching (IM) Accuracy: percentages of samples correctly classified by the integrated system.

To assess the robustness of PAD systems under biometric recognition constraints (Challenge 1), we conducted comparision trials on three devices: GreenBit, Dermalog, and Huawei. For each device, we generated comparison scores between a fingerprint template and (i) a bona fide image from the same user and finger (mated), (ii) a presentation attack from the same user and finger (attack), or (iii) a bona fide image from a different user (non-mated). The number of mated comparisons varies depending on available bona fide samples per sensor (e.g., 3110 for GreenBit, 3070 for Dermalog, and 1240 for Huawei). We generated twice as many non-mated and attack comparisons for each mated comparison set. No comparison trials were conducted on the remaining devices due to the limited number of bona fide samples.

Challenge 2 focused on assessing the trade-off between efficiency and discriminative power of the feature representations extracted by each system. Particular emphasis was placed on the computational speed and the compactness of the resulting feature vectors. To ensure consistency across submissions, all evaluations were conducted on standardized hardware platforms: either a Linux (Ubuntu 18.04.1) or Windows 10 Pro desktop equipped with an Intel® Core™ i9-9900K CPU (3.60GHz), 64 GB DDR4 RAM (2.933 MHz), and dual NVIDIA® GeForce® RTX 2080 Ti GPUs (11 GB each). The final score for each algorithm was computed as the arithmetic mean of three normalized components: feature extraction time (s_t), feature vector size (s_d), and PAD accuracy (s_a), defined as follows:

$$s_t = 1 - \frac{T - T_{\text{min}}}{T_{\text{max}} - T_{\text{min}}};$$

$$s_d = 1 - \frac{D - D_{\text{min}}}{D_{\text{max}} - D_{\text{min}}}; \quad s_a = \frac{A}{100}$$

where T is the average time per comparison (in seconds), D is the feature dimensionality, and A is the PAD accuracy (in percent). For LivDet2025, we set $T_{\text{min}} = 0.001$, $T_{\text{max}} = 1$, $D_{\text{min}} = 1$, and $D_{\text{max}} = 512$. The overall score was then computed as

$$s_f = \frac{s_t + s_d + s_a}{3}$$

2.4. Adversarial Threat Model

To assess adversarial robustness, we considered a black-box threat model in which the attacker has no knowledge of the internals of the targeted PAD system but assumes it is based on deep neural networks. Given the practical constraints of real-world deployments, such a setting captures realistic attack scenarios where system internals are proprietary or obscured. All submitted PADs

rely on convolutional or transformer-based feature extractors 3: *Dermalog*'s system combines MobileNetV2 with RNN-based fusion of fingerprint minutiae patches; *JIIOV* adopts a transformer-based architecture over keypoint features; *UNIFESP* leverages an hybrid approach based on frozen CNNs for feature extraction. Under the black-box assumption, adversarial samples were generated via a surrogate-based transfer approach, following the strategy proposed in [6]. Let f_{target} be the decision function of the unknown PAD to be attacked. The adversary trains a surrogate model $f_{\text{surr}} \approx f_{\text{target}}$ using an auxiliary dataset and then crafts adversarial examples:

$$\mathbf{x}' = \mathbf{x} + \boldsymbol{\delta}, \quad \text{with} \quad \|\boldsymbol{\delta}\|_\infty \leq \epsilon, \tag{1}$$

such that:

$$f_{\text{surr}}(\mathbf{x}') \geq \tau \quad \text{and} \quad f_{\text{target}}(\mathbf{x}') \geq \tau, \tag{2}$$

where $\tau = 50$ is the fixed liveness threshold. The goal is to shift the prediction of previously rejected (PA) samples into the bona-fide region by leveraging the transferability of adversarial perturbations across models. In our setup, the surrogate model is a VGG19 network pre-trained on ImageNet and subsequently fine-tuned on fingerprint images from the LivDet 2015 dataset. Only images initially classified as PA by the target PADs were selected for attack. Perturbations were generated using the Auto-PGD (APGD) algorithm [11], a momentum-based iterative method known for its effectiveness under both white-box and black-box conditions. The attack operates under an ℓ_∞ constraint and integrates adaptive step sizes and multiple restarts to enhance convergence and overcome potential gradient obfuscation.

3. Results

The results of the contact-based PAD algorithms submitted to LivDet2025 are reported in Tables 4–10. All participants provided complete entries for both Challenge 1 and Challenge 2. While limited in number, the submissions offer instructive contrasts regarding trade-offs between PAD accuracy, biometric verification, and computational complexity. At a high level, the three PAD systems exhibit markedly different trade-offs between security and usability. As shown in Table 4, *UNIFESP* prioritizes presentation attack rejection, yielding the lowest APCER (16.4%) and IAPAR (0.22%) across all systems. However, this robustness comes at a significant usability cost: the integrated system fails to perform almost any meaningful mated comparison, as evidenced by a FNMR of 99.7%, suggesting a substantial mismatch between the feature representations or calibration strategies adopted by the PAD and the recognition components. On the other hand, *Derm_C* shows the opposite behavior: bona fide samples are consistently above the threshold (BPCER = 1.0%, FNMR = 18.8%), but PA scores largely overlap, resulting in an APCER of 96.2% and IAPAR of 50.7%. *jiiov* achieves the most effective score separation, with moderate APCER (27.3%) and low BPCER (1.3%), yielding the highest IM accuracy (89.5%) along with a reduced attack success rate (IAPAR = 16.5%).

Analyzing performance across datasets (Table 6), results on the consensual subsets (DLCC, GBCC) tend to be stronger overall, while the ScreenSpoof subsets (DLSS, GBSS) are generally more challenging. In most cases, APCER increases, reflecting the added difficulty posed by the use of an unconventional PAI creation technique. This trend is mirrored in Fig. 2, which shows per-material APCER grouped by sensor: some materials, such as the GLSpro and GSP400n used for the realization of PAI with the ScreenSpoof technique, lead to high errors, often above 50%. This shows that the problem of generalization of PADs to unknown attacks is still an open problem [13].

Challenge 2 results (Table 5) highlight key trade-offs between accuracy, latency, and template size in deployment-oriented scenarios. The *jiiov* system attains the highest composite score (0.85), combining competitive PAD accuracy (83.3%) with low computational latency (27 ms/image) and a compact embedding size (128 floats). *UNIFESP* achieves slightly higher accuracy (87.3%) at the expense of a fivefold increase in processing time (152 ms/image), while *Derm_C* exhibits the least favorable trade-off: despite a comparable accuracy (84.9%), its inference time exceeds 560 ms per image. These results suggest that the primary source of latency lies in the pre-embedding stages (e.g., segmentation, quality enhancement), rather than in feature extraction or representation. Notably, all three systems rely on lightweight templates ($\leq$ 138), reflecting a general trend toward compact descriptors in modern PAD systems compared to earlier editions [4, 18].

The contactless track saw only one submission, *Derm_CL*, whose performance varied sharply across acquisition devices (Table 8). While near-perfect on iPhone and Galaxy samples (PAD ACC $\geq$ 95%), its accuracy plummeted on the Huawei P20 Pro (56%), revealing a strong dependency on device-specific imaging characteristics.

The contactless track received a single submission, *Derm_CL*, which achieved high accuracy on iPhone and Galaxy samples (PAD ACC $\geq$ 95%), but dropped to 56% on the Huawei P20 Pro, indicating limited generalization. According to the accompanying technical report, the training data included sensors also present in the test set (see Table 3, effectively making the evaluation intra-dataset. The substantial performance drop on the unseen Huawei device highlights the need for per-device calibration or meta-learning strategies capable of adapting PAD behavior to novel acquisition conditions without prior knowledge of the sensor or attack type.

The results of Challenge 3 are reported in Table 10 in

Table 4. Challenge 1: Integrated and PAD overall results (%) for contact-based detectors.

Detector	FMR	FNMR	IAPAR	IM Acc	BPCER	APCER	PAD ACC
UNIFESP	0.14	99.69	0.22	79.92	7.46	16.41	88.96
Derm_C	0.28	18.84	50.69	75.84	1.02	96.18	60.92
jiiov	0.11	19.16	16.47	89.54	1.32	27.32	88.28

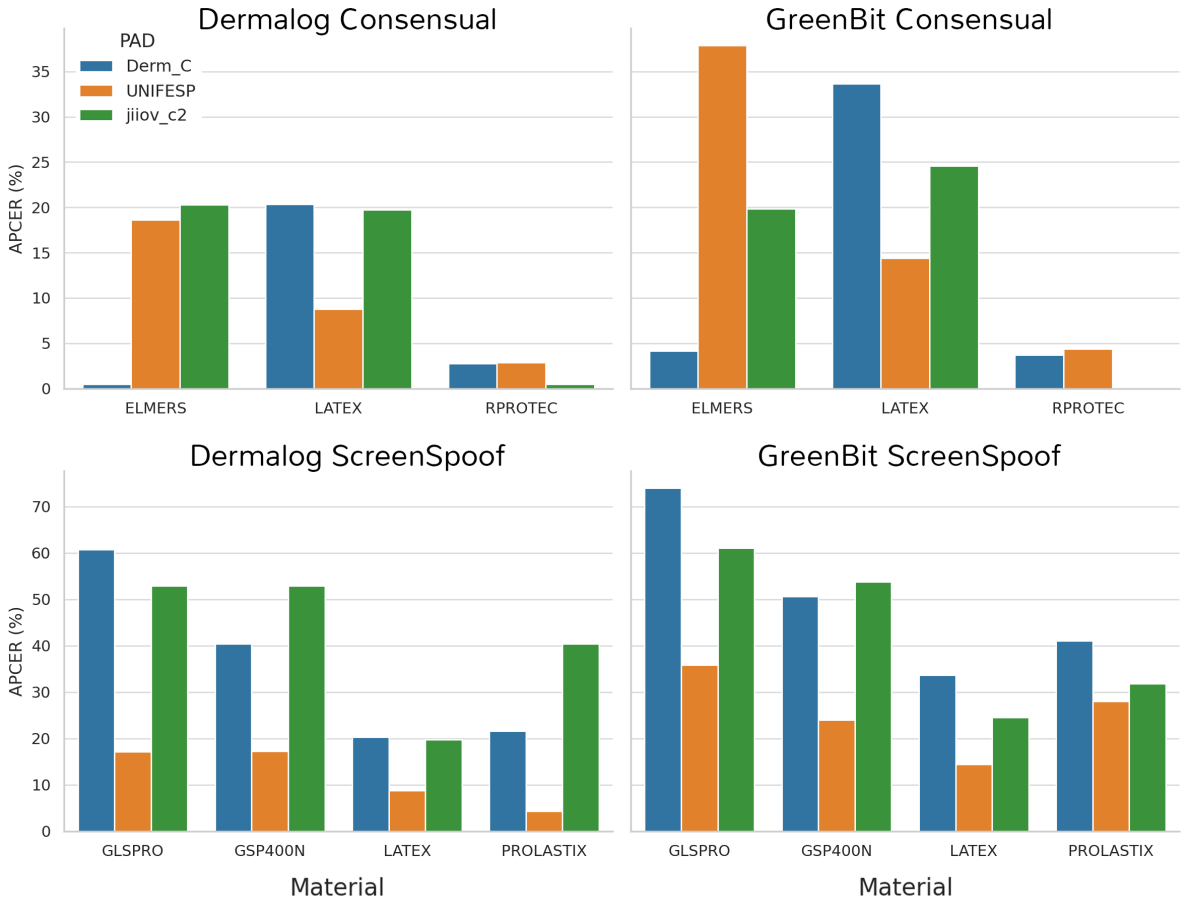

Figure 2. Material-specific APCER across PAD systems for each sensor. Top row: results on contact-consensual (CC) subsets acquired with Dermalog (left) and GreenBit (right) sensors. Bottom row: results on ScreenSpoof (SS) subsets for the same sensors.

Table 5. Challenge 2: PAD overall results for contact-based and contactless detectors.

Algorithm	time/img [ms]	feat.size	ACC[%]	Score
UNIFESP	151.70	138	**87.28**	0.82
Derm_C	560.20	**128**	84.94	0.68
jiiov	**27.41**	**128**	83.27	**0.85**
Derm_CL	37.23	32	87.26	–

terms of APCER and FTX relative to the attacked samples only. Table 9 reports the number of successful adversarial attacks for each material with respect to the total number of original samples.

The results reveal three distinct behaviors. *Derm_C* shows signiticant increases in FTX (up to 88%) but apparent drops in APCER (e.g., from 51.7% to 0.00% on GLS SS). This improvement is due to the adversarial noise disrupting the feature extraction process, leading to many samples being rejected early (i.e., no decision taken), thereby decreasing APCER and increasing FTX. We note that this is not an entirely negative result for PAD if we consider that an FTX corresponds to an immediate rejection. *jiiov*, built on a ViT backbone, preserves stable feature extraction (FTX = 0.00%), but experiences a clear rise in APCER across most materials (+20–40 APCER), showing that the adversarial attack was successful. *UNIFESP*, by contrast, proves the most robust overall: APCER remains low ($\leq 6\%$) and FTX is unaffected. This suggests that the frozen CNN layers and the traditional classifier used in UNIFESP are poorly aligned with VGG-19 gradients, limiting the transferability of attacks. It is interesting to analyze the effects of adversar-

Table 6. Challenge 1: Integrated and PAD performance (%) across datasets and attack types for each competitor.

Dataset	Detector	FMR	FNMR	IAPAR	IM Acc	BPCER	APCER	PAD ACC	FTX
DLCC	UNIFESP	0.36	99.45	0.07	79.94	12.5	8.5	89.1	0
	Derm_C	0.20	18.44	60.20	72.15	1.17	98.18	60.03	0.18
	jiiov	0.10	18.18	5.00	94.32	0.93	6.99	96.64	0
DLSS	UNIFESP	0.20	99.29	0.33	79.93	12.97	14.11	86.57	0
	Derm_C	0.30	19.13	40.17	79.99	1.58	95.31	60.93	1.24
	jiiov	0.08	18.01	25.92	86.00	0.76	46.48	80.96	0
GBCC	UNIFESP	0.00	100.00	0.13	79.95	2.29	14.84	92.69	0
	Derm_C	0.35	19.23	59.42	72.24	0.59	98.26	60.34	0.36
	jiiov	0.10	20.68	4.45	94.05	1.63	6.48	96.43	0
GBSS	UNIFESP	0.00	100.00	0.35	79.86	2.07	28.2	87.48	0
	Derm_C	0.25	18.55	42.98	78.99	0.75	92.98	62.36	2.44
	jiiov	0.15	19.78	30.49	83.79	1.96	49.32	79.10	0
HWCL	Derm_CL	0.00	97.66	0.32	80.34	95.32	4.76	40.90	0

Table 7. Challenge 2: PAD performance (%) of contactless systems across different smartphone models.

Algorithm	Huawei			iPhone 7			iPhone X			Galaxy S7			Overall PAD ACC
	BPCER	APCER	PAD ACC	BPCER	APCER	PAD ACC	BPCER	APCER	PAD ACC	BPCER	APCER	PAD ACC	
Derm_CL	93.47	0.12	56.46	1.35	3.41	96.81	0.84	4.37	95.95	0.00	0.55	99.83	87.26

Table 8. Challenge 2: PAD performance of contact-based systems across datasets and attack types for each competitor.

Algorithm	Dermalog							GreenBit							Overall PAD Acc.
	BPCER	Consensual			ScreenSpoof			BPCER	Consensual			ScreenSpoof			
		APCER	PAD Acc	FTX	APCER	PAD Acc	FTX		APCER	PAD Acc	FTX	APCER	PAD Acc	FTX	
UNIFESP	13.31	8.19	89.40	0.00	13.24	86.73	0.00	2.17	15.40	90.81	0.00	28.07	82.20	0.02	87.29
Derm_C	1.83	1.24	98.48	0.86	40.27	76.06	6.41	0.86	3.25	97.90	1.90	57.53	67.34	9.59	84.95
jiiov	0.86	7.12	95.83	0.00	46.21	71.83	0.00	1.83	6.61	95.63	0.00	48.84	69.81	0.00	83.28

Table 9. Number of successful adversarial attacks over the total number of original samples, for each PAI material.

	latex CC	elmers CC	gls SS	gsp SS	latex SS	rprolast SS	rprotec CC
Attacked	604/680	376/660	441/650	476/672	525/525	493/670	564/641

Table 10. Challenge 3: APCER and FTX values (%) for original and adversarial samples across diferent materials. Adversarial attacks were performed only on samples from the GreenBit subset. APCER* and FTX* are calculated on samples attacked with adversarial perturbations (see Tab. 9).

Greenbit material		Latex CC	Elmers CC	GLS SS	GSP SS	Latex SS	RProlast SS	RProtec CC
Algorithm	**Metric**	**original/adv**	**original/adv**	**original/adv**	**original/adv**	**original/adv**	**original/adv**	**original/adv**
Derm_C	**APCER***	1.32 / 0.17	1.06 / 0.00	51.70 / 0.00	31.30 / 0.00	0.00 / 0.00	23.94 / 0.00	3.55 / 0.00
	FTX*	0.00 / 70.86	6.65 / 88.03	19.05 / 33.11	18.91 / 33.61	57.52 / 57.52	17.44 / 35.29	2.66 / 80.50
jiiov	**APCER***	8.44 / 0.00	18.62 / 32.71	58.73 / 71.88	49.16 / 78.57	57.90 / 57.90	30.63 / 78.09	0.00 / 0.00
	FTX*	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00
UNIFESP	**APCER***	4.47 / 0.66	28.19 / 6.12	29.71 / 3.40	16.81 / 1.47	1.33 / 1.33	26.77 / 1.22	2.48 / 1.60
	FTX*	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00

ial attacks on individual materials. Latex-based PAs remain hard to perturb, showing minimal APCER variation pre/post attack. Elmers-based PAs highlight system disparity: while UNIFESP resists transfer (from 28.2% to 6.1%), *jiiov* shows a marked increase (from 18.6% to 32.7%). ScreenSpoof materials (GSP, RProlast) emerge as the most perturbable, with *jiiov* seeing APCER $\geq$ 75%. The observed behaviors suggest that adversarial transferability is closely tied to architectural choices. ViTs with high-dimensional embeddings are more gradient-aligned with VGG-based surrogates, while early feature extractors like those in *Derm_C* are vulnerable to low-level perturbations that block the entire pipeline.

A composite view across the three challenges leads to several conclusions. First, threshold calibration alone (e.g., UNIFESP) can swing a system from high PAD accuracy to operational failure. Second, APCER and FTX must be interpreted jointly in adversarial settings: an alteration of the fingerprint characteristics that lead to an automatic rejection depending on the analysis carried out by the PAD. Third, the success of adversarial attacks is strictly related to the type of architecture used: PADs based on widely adopted, open-source models are particularly vulnerable and should consider incorporating adversarial training or robust blocking

strategies. It is also important to note that the attacks performed in this evaluation were relatively naive and entirely black-box. More sophisticated approaches, such as employing alternative adversarial techniques or combining multiple white-box classifiers, could potentially result in significantly greater degradation of PAD performance.

4. Discussion and conclusions

The ninth edition of the Fingerprint Liveness Detection Competition (LivDet2025) introduced several significant innovations, including the use of contactless fingerprints, an adversarial robustness track, and the removal of standardized training data. This last point simulated a realistic "worst-case" deployment scenario, requiring participants to rely entirely on publicly available datasets, thereby emphasizing true generalization capability.

Results reveal that, while progress is evident, the generalization problem remains unresolved. In particular, detection performance degrades on specific sensors and novel fabrication techniques, especially in the ScreenSpoof subsets. Nonetheless, some solutions demonstrated promising robustness, even in the face of black-box adversarial attacks.

A clear trade-off emerged between PAD accuracy, operational usability, and computational efficiency. Some systems prioritized robustness but suffered in recognition performance, while others favored speed and simplicity at the expense of attack resilience.

Acknowledgements

This research work has been partially funded by project FAIR (PE00000013) under the NRRP MUR program funded by the EU–NGEU (CUP: J23C24000090007) and by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] ISO/IEC 2382-37:2022 Information technology — Vocabulary — Part 37: Biometrics, 2022.

[2] ISO/IEC 30107-3:2023 Information technology — Biometric presentation attack detection — Part 3: Testing and reporting, 2023.

[3] D. S. Ametefe, S. S. Sarnin, D. M. Ali, G. D. Ametefe, D. John, and N. Hussin. Advancements and challenges in fingerprint presentation attack detection: a systematic literature review. *Neural Computing and Applications*, pages 1–23, 2024.

[4] R. Casula, M. Micheletto, G. Orrù, R. Delussu, S. Concas, A. Panzino, and G. L. Marcialis. Livdet 2021 fingerprint liveness detection competition-into the unknown. In *2021 IEEE international joint conference on biometrics (IJCB)*, pages 1–6. IEEE, 2021.

[5] R. Casula, M. Micheletto, G. Orrú, G. L. Marcialis, and F. Roli. Towards realistic fingerprint presentation attacks: The screenspoof method. *Pattern Recognition Letters*, 2022.

[6] R. Casula, G. Orrù, S. Marrone, U. Gagliardini, G. L. Marcialis, and C. Sansone. Realistic fingerprint presentation attacks based on an adversarial approach. *IEEE Transactions on Information Forensics and Security*, 19:863–877, 2023.

[7] I. Chingovska, A. Anjos, and S. Marcel. Anti-spoofing in action: Joint operation with a verification system. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 98–104, 2013.

[8] A. M. Chowdhury and M. H. Imtiaz. Contactless fingerprint recognition using deep learning—a systematic review. *Journal of Cybersecurity and Privacy*, 2(3):714–730, 2022.

[9] R. C. Contreras, W. Molta, L. Feng, M. S. Nagahara, M. Nappi, P. Campisi, and A. de Albuquerque Araújo. A new multi-filter framework with statistical dense sift descriptor for spoofing detection in fingerprint authentication systems. In *International Conference on Artificial Intelligence and Soft Computing (ICAISC)*, pages 442–455, Cham, 2021. Springer.

[10] R. C. Contreras, W. Molta, M. S. Nagahara, L. Feng, X. Liu, M. Nappi, P. Campisi, and A. de Albuquerque Araújo. A new multi-filter framework for texture image representation improvement using set of pattern descriptors to fingerprint liveness detection. *IEEE Access*, 10:117681–117706, 2022.

[11] F. Croce and M. Hein. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. In *International conference on machine learning*, pages 2206–2216. PMLR, 2020.

[12] J. Galbally-Herrero, J. Fierrez-Aguilar, J. Rodriguez-Gonzalez, F. Alonso-Fernandez, J. Ortega-Garcia, and M. Tapiador. On the vulnerability of fingerprint verification systems to fake fingerprints attacks. In *Proceedings 40th Annual 2006 International Carnahan Conference on Security Technology*, pages 130–136. IEEE, 2006.

[13] A. K. Jain, D. Deb, and J. J. Engelsma. Biometrics: Trust, but verify. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3):303–323, 2022.

[14] J. Kolberg, M. Gomez-Barrero, and C. Busch. On the generalisation capabilities of fingerprint presentation attack detection methods in the short wave infrared domain. *IET Biometrics*, 10(4):359–373, 2021.

[15] J. Kolberg, J. Priesnitz, C. Rathgeb, and C. Busch. Colfispoof: A new database for contactless fingerprint presentation attack detection research. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 653–661, 2023.

[16] H. Li and R. Ramachandra. Deep learning based fingerprint presentation attack detection: A comprehensive survey. *arXiv preprint arXiv:2305.17522*, 2023.

[17] D. Maltoni, D. Maio, A. K. Jain, J. Feng, et al. Handbook of fingerprint recognition. 2022.

[18] M. Micheletto, R. Casula, G. Orrù, S. Carta, S. Concas, S. M. La Cava, J. Fierrez, and G. L. Marcialis. Livdet2023-

fingerprint liveness detection competition: advancing generalization. In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2023.

[19] M. Micheletto, G. Orrù, R. Casula, D. Yambay, G. L. Marcialis, and S. Schuckers. Review of the fingerprint liveness detection (livdet) competition series: from 2009 to 2021. *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pages 57–76, 2023.

[20] M. Micheletto, G. Orrù, L. Ghiani, and G. L. Marcialis. Improving fingerprint presentation attack detection by an approach integrated into the personal verification stage. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2025.

[21] K. Okereafor, I. Ekong, I. O. Markson, K. Enwere, et al. Fingerprint biometric system hygiene and the risk of covid-19 transmission. *JMIR Biomedical Engineering*, 5(1):e19623, 2020.

[22] S. Purnapatra, H. Rezaie, B. Jawade, Y. Liu, Y. Pan, L. Brosell, M. R. Sumi, L. Igene, A. Dimarco, S. Setlur, et al. Liveness detection competition-noncontact-based fingerprint algorithms and systems (livdet-2023 noncontact fingerprint). In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2023.

[23] J. M. Singh, A. Madhun, G. Li, and R. Ramachandra. A survey on unknown presentation attack detection for fingerprint. In *Intelligent Technologies and Applications: Third International Conference, INTAP 2020, Grimstad, Norway, September 28–30, 2020, Revised Selected Papers 3*, pages 189–202. Springer, 2021.