Explainable Presentation Attack Detection of Digital Fingerprints

Augustin Diers, Christophe Rosenberger

Explainable Presentation Attack Detection of Digital Fingerprints

Augustin Diers and Christophe Rosenberger
Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR6072, F-14000 Caen, France
augustin.diers@etu.unicaen.fr, christophe.rosenberger@ensicaen.fr

***Abstract*—Biometric systems are used in our daily life but are subject to attacks to bypass them as a security solution. Presentation attacks in digital fingerprints occur when an imposter tries to use a fake sample at the acquisition step to impersonate another individual or not to be identified. Providing an explanation for the operator (who is not an expert in biometrics) could be of great interest for many applications (border control, physical access control). In this paper, we propose a fingerprint presentation attack detection method with explainability feedback that can be understood by any user. The experiments has been realized on the Fingerprint Liveness Detection Competition (LivDet) dataset in 2015 and contains more than 58,000 bona fide and attack fingerprint images. The proposed method reaches an accuracy rate of 95.7% on LivDet2015 with feedback that can be understood by any user.**

***Index Terms*—Biometrics, Presentation attack detection, Digital fingerprints, Explainability.**

I. Introduction

Biometrics has for objective to automatically identify an user or verify its identity by using morphological or behavioral characteristics. Nowadays, digital fingerprint is used as one of the most secure and reliable biometric modality for user authentication. Fingerprints tend to replace passwords in applications requiring user authentication. This massive use of fingerprints as a security solution has therefore led to the appearance and multiplication of attacks on such systems. For example, a Chinese woman expelled from Japan in 2009 managed to reintroduce herself into the territory by replacing her left fingerprints with those of the right hand. The border control system failed when searching the list of people expelled and prohibited from Japanese territory. It is therefore important to add a fourth function to any biometric system that is called Presentation Attack Detection (PAD) or Anti-Spoofing [1]. The first three functions being enrollment, authentication and identification as described in [2]. Presentation attacks occur on the biometric sensor and the imposter attempts to forge the biometric template of another person or create a new biometric fingerprint template in order to access confidential information to which he/she has no right.

Cooperative and non-cooperative methods are the two ways used by impostors to forge fake fingerprints called presentation attack instrument (PAI). In cooperative methods, the impostor collaborates with the individual whose identity he/she wants to impersonate in order to obtain a perfect mold of his/her fingerprint (for time attendance as for example). In non-cooperative methods, the impostor tries to forge the fingerprint template of an individual without his/her consent from different sources (latent, cadavers, synthetic). In both cases, the impostors use gummy materials (Latex, silicone, playdoh…) to generate the shape of the fingerprint to impersonate. Hardware and software solutions are the two ways to counter the presentation attacks proposed in the literature [3]–[5]. In [6], the authors proposed different hardware solutions for PAD. These solutions need to integrate specific components on the sensor to measure the ridges distortion, elasticity, temperature and body conductivity. Software solutions are subdivided into two groups: dynamic and static methods [7]. The dynamic PAD methods use the fingerprint features that are expected to vary on a video stream on the sensor. To implement this solution, two or more images of the fingerprint are acquired between two very short instants of the capture (from 0 to 5 seconds) as shown in [8]. Nowadays, most static PAD methods are based deep learning such as Convolutional Neural Networks (CNN) or transformers [9].

An important lack of these black box solutions is the difficulty to trust the decision result without any understandable feedback even performance results are usually very high. The main contribution of this paper is to propose different solutions to enhance the explainability in biometric security. It is necessary to better trust the decision provided by a biometric system or PAD module and help to identify attacks by non-expert operators. The paper is organized as follows. In section II, we present a state of the art on Explainable Artificial Intelligence (XAI) in order to provide useful feedback by a PAD. Section III is dedicated to the proposed method. Section IV and V concern respectively the experimental protocol and results. We conclude and give some perspectives on section VI.

II. State of the art

We focus in this state of the art on XAI solutions that could be used for PAD systems. Opaque systems are often untrustworthy [20], even when an explanation is attempted. Black-box systems decrease user trust, particularly after an incorrect prediction, as they obscure the system reasoning process. Educating users does not sufficiently address the lack of explainability. The adoption of such systems relies on the ability to translate the model process into a shared language between the system and the user. Researchers and engineers

TABLE I
MANY INTERESTING XAI SOLUTIONS FOR BIOMETRICS.

Article	Description
Using of Grad-CAM to detect presentation attack. [10]	presentation attacks for iris
Embedding deep networks into visual explanations [11]	Novel Explanation Neural Network (XNN) called Sparse Reconstruction Autoencoder (SRAE). It maps the output embedding into an explanation space capable of retaining the prediction power of the original feature embedding.
Bayes-trex: a bayesian sampling approach to model transparency by example [12]	Display the boundaries between the classes through ambiguous samples by finding in-distribution examples with specified prediction confidence.
Explaining explanations: An overview of interpretability of machine learning [13]	Textual explanations as part of their training process.
SHapley Additive exPlanation (SHAP) is a game-theoric approach to explain ML predictions [14].	Features represented as players in a coalition game. The payoff is the Shapley value, an additive measure of importance.
Class activation maps (CAMs) which are specific to Convolutional Neural Network CNNs) [15]	Global average pooling applied to the final convolutional feature map, before the output layer. They are then used as the input features of a fully connected layer and output through a loss function. We can then, by projecting the weights back to the previous convolutional layer, create an image with the areas of greater influence over the CNN decisions highlighted per class and visible through a heatmap representation.
Interpretable Local Surrogates [16]	Method that aims to replace the decision function by a local surrogate model that is structured in a way that it is self-explanatory (like a linear model).
Occlusion Analysis [17]	Particular type of perturbation analysis where we repeatedly test the effect on the neural network output, of occluding patches or individual features in the input image. A heatmap can be generated from the scores.
Integrated Gradients [18]	Integrating the gradient $\nabla f(x)$ along some trajectory in input space connecting some root point $\tilde{x}$ to the data point x. It is well-suited to explain functions that have multiple scales. To be implemented, integrated gradient must be discretized, it is then approximated by a sequence of data points $x^{(1)}, ..., x^{(N)}$
Layer-Wise Relevance Propagation (LRP) [19]	The layered structure of the neural network is operated in an iterative manner until we reach the output layer. Then, a reverse propagation is applied, where the output score is redistributed, layer after layer.

have attempted to deploy biometric systems at scale in real-world scenarios. Despite efforts to build trust in these systems and demonstrate their unbiased and accurate behavior, failures have occurred [21]. A critical step to address this issue is to provide users with explanations of algorithmic decisions. This approach can encourage users to trust the results, alert them to potential attacks, and help identify unexpected behaviors or biases. Additionally, providing counterexamples and introducing contrastiveness can help in user understanding.

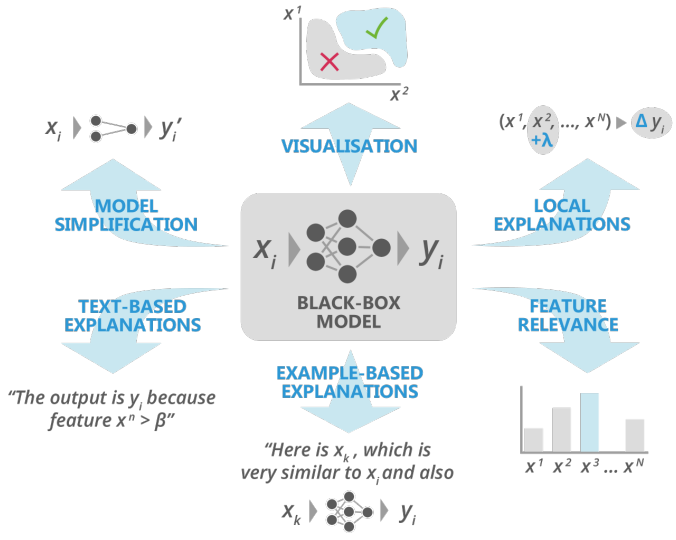

Fig. 1. Different approaches for the explainability of decisions given a biometric system [22].

We can divide the explainability into three smaller subtopics [22]: pre-models techniques, in-models techniques and post-model techniques (or post hoc).

- "Pre-model techniques" (or ante-hoc) focuses on data knowledge and understanding. It is quite relevant to understand some bias in the data and can increase confidence in the posterior decisions and explanations.
- "in-model techniques" focus on the direct integration of interpretability into the model through the constraint applied to the training process.
- "post-model techniques" (or post hoc) refer to methods used to interpret and understand the decisions made by an artificial intelligence model after it has already made its predictions or decisions.

The high requirement for accuracy and other performance metrics justify that the usage of post hoc methods is much more convenient and popular. Different explanation methods lead to different qualities of explanation (see Table I). Evaluating explanations remains a difficult task. Machine learning models are usually evaluated by the utility of their decision behavior. Transposed to the domain of explanation, it would require to define what is the ultimate target and assess by how much the use of explanation increases its performance on the target task, compared to not using it. In the next section, we present the proposed method producing explanations for the PAD system.

III. PROPOSED METHOD

The objective of the proposed method is to build a classical machine learning workflow for presentation attack detection on

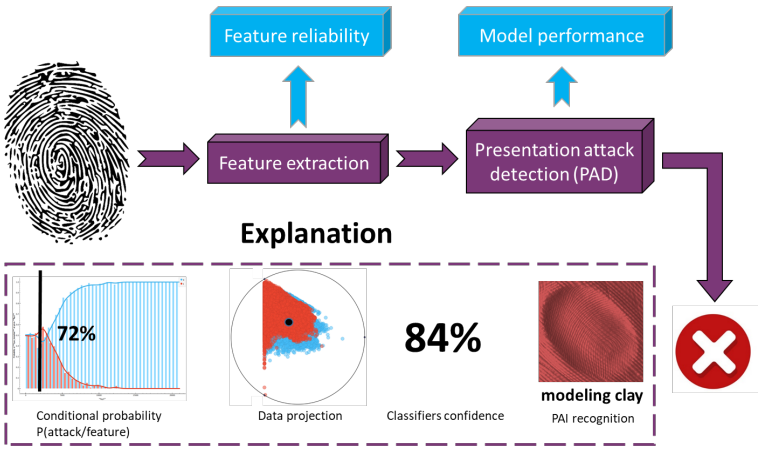

Fig. 2. Proposed methodology for explaining decisions of a presentation attack detection method of fingerprints.

fingerprints by adding intelligible explanations for an operator. For this task, we preliminary analyze the reliability of used features (see Figure 2). In the following, we detail first the used features and then the machine learning workflow.

A. Feature extraction

We use in this work hand-crafted features that better help to explain resulting decisions. As we work on digital fingerprints, minutiae are considered in this study [23]. The minutia indicates the different discontinuities of a local ridge. There are two types: the ending minutiae (ridge ending) and bifurcations. The minutiae representation is stable, robust, unique and has a non uniform distribution. A minutia m can be described by $m(x, y, t, \theta, q, dst_1, nb\_cr_1, dst_2, nb\_cr_2, dst_3, nb\_cr_3)$ where x-axes y-axes are coordinates, t its type, θ its orientation and then the quality index q associated with the fingerprint. dst_1, nb_cr_1, represent respectively the distance which separates the minutia from its nearest neighbor and the number of ridges which separate them. The indices 2 and 3 represent the same measures with the second and third neighbors.

For a fingerprint I, we consider each of the properties of a minutia, except coordinates (x, y), as variables that can be used for PAD. Therefore, we calculate the four statistical estimators we described above ($\bar{x}$, $E(X)$, S and K) for the set of minutiae of I. For example, $\bar{q}$ indicates the overall local quality level of the minutiae detected on the fingerprint, and then $E(q)$ determines the deviation of the individual quality indices of each minutia compared to the central tendency $\bar{q}$. On fake fingerprints, $E(q)$ seems to be small because the fingerprint image has a strong homogeneity and low variations unlike live fingerprints for which the values are more random. The same observations are made on the variable θ which indicates the local variation of the directions of the minutiae and demonstrates the homogeneity on the fake fingerprints unlike the real ones. The statistical estimators that we use, explore the minutiae variations and we build 36 features from this local fingerprint expertise.

We believe that the quality of a fingerprint is a crucial element for presentation attack detection. Whatever the effort made by an impostor, the homogeneous nature of the ridges introduced in a fake fingerprint constitutes a failure compared to a live fingerprint. For this reason, we use the quality map of fingerprint. The map indicates for each area of the image the associated quality value. The overall quality of the live fingerprint is better than the one of a fake fingerprint. In addition, the values being between 0 and 5 (5 for the highest quality), we use this map to build the 7 following features ; the global quality measures (Total sum); as well as the relative frequencies of each local quality value of the image from 0 to 5. Likewise, manufacturing faults in a fake fingerprint and sometimes the difficulty of putting it uniformly on the sensor leads to the appearance of small holes (blank area) in the image of the fingerprint. And so, a fake fingerprint image has more blank areas than a live fingerprint image. The hole frequencies are extracted from the Low Flow Map and allows to identify blank areas. As we explained, fake fingerprints usually have a high homogeneity; so, we use the Direction Map to locate globally on the image the change of frequency direction as well as the errors occurred by the extractor. The areas of high curvature of the fingerprint are identified with the extractor from the High curvature Map. This indicates the singular points of fingerprint know as delta, loop, arch, core. Thus, there are limited number of singular points for a live fingerprint and a little more in the fake. In total, we build 49 new features based on the specific fingerprint expertise. These features are frequency estimators and statistical descriptors resulting from the fingerprint analysis. We therefore propose the extraction of these features which we concatenate with the static texture features proposed in the state of the art in order to build a more robust representation. We have chosen to combine the semantic features with LBP ones related to texture analysis.

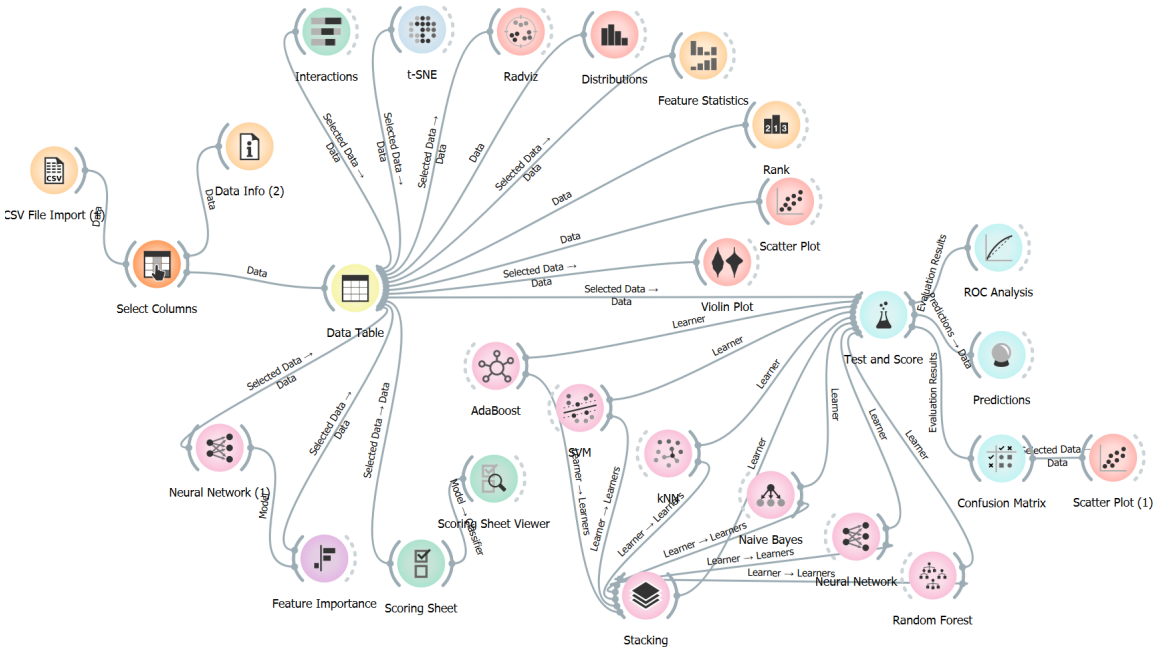

Fig. 3. Workflow for the presentation attack detection method with Orange Data Mining.

B. Machine learning workflow

The machine learning workflow is described in Figure 3. It has been generated with the Orange Data Mining software allowing to define workflows in data science [24]. In this work, we used 6 machine learning models among Adaboost, SVM (RBF), K nearest neighbors, random forest, naive Bayes and neural network. A meta model is generated from previous ones. For the training and testing, we adopt a cross validation process with 5 folds.

IV. Experimental protocol

A. Dataset

The most used databases for PAD evaluation are LivDet. Since 2009, the LivDet competition started and many contestants proposed their PAD model solutions for this benchmark. All solutions are evaluated and the benchmark of PAD models are provided at the end of the competition. We use in this work the LivDet15 dataset [25] in reference to the competition in 2015. Figure 4 shows the tree structure of the LivDet2015 database. In this tree structure, the terms Alive, Spoof and spoof_MaterialName should be noted, which respectively mean alive fingerprint, fake fingerprint and then fake fingerprint with the material used for its manufacture. Biometrika, Green Bit, Digital persona and Crossmatch sensors are used for fingerprint acquisition. The dataset is composed of 58583 samples (30471 bona fide, 28112 attacks).

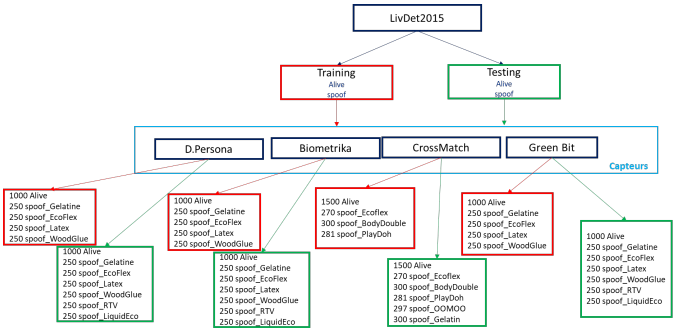

Fig. 4. LivDet2015 database tree [25].

B. Evaluation metrics

Indeed, after building the PAD model, we need to define evaluation metrics. In the case of presentation attack detection performance measurement, we consider three metrics:

- APCER (Attack presentation classification error rate): The percentage ratio at which presentation attack examples are misidentified as a bona fide example. A higher value indicates higher security vulnerability.
- BPCER (Bona fide presentation classification error rate): The percentage ratio at which bona fide examples are misidentified as a presentation attack example. A higher value indicates higher user friction.
- CA (Classification Accuracy): It measures the correct recognition of bona fide and attack examples.

V. Experimental results

A. Performance results

First, we present the performance of the machine learning workflow with the proposed hand-crafted features. Figure 5 shows the classification accuracy for each model. The stacking model achieves a high performance with a CA value of 95.7%. Figure 6 presents the confusion matrix associated to this result.

In this case, we obtain a similar APCER and BPCER values of 2.1% showing the benefit of the proposed method.

Model	AUC	CA	F1	Prec	Recall	MCC
SVM	0.580	0.541	0.530	0.554	0.541	0.101
Naive Bayes	0.688	0.630	0.629	0.629	0.630	0.258
AdaBoost	0.829	0.830	0.830	0.830	0.830	0.659
kNN	0.914	0.867	0.867	0.867	0.867	0.734
Random Forest	0.967	0.906	0.905	0.906	0.906	0.811
Neural Network	0.987	0.947	0.947	0.947	0.947	0.893
Stack	0.991	0.957	0.957	0.957	0.957	0.914

Fig. 5. PAD performance results for each learning model.

		Predicted 0	Predicted 1	Σ
Actual	0	29246	1225	30471
	1	1286	26826	28112
	Σ	30532	28051	58583

Fig. 6. Confusion matrix for the presentation attack detection method (0: bona fide, 1: attack).

B. Explainability

We showed in the previous section that the proposed method achieves a very good efficiency. Even if APCER et BPCER values are low, errors can occur. In sensitive applications such as border control, we might be interested to give an understandable explanation to the operator. As illustration, we used a specific sample corresponding to an attack (presentation attack instrument) in Figure 7.

We propose different explanations or feedback to the operator:

- Single feature analysis: Before deploying the PAD system, a preliminary step can be done to identify reliable features. We can plot the distribution of the values of each feature for all classes (bona fide and attacks). Figure 8 shows an illustration of these distributions (one LBP feature). We can estimate given the feature value (represented by a dark line) the conditional probability value the sample corresponds to an attack. In the example of Figure 8, we can see clearly that the higher is the feature the more confidence the sample is bona fide.

Fig. 7. Attack sample used for illustration

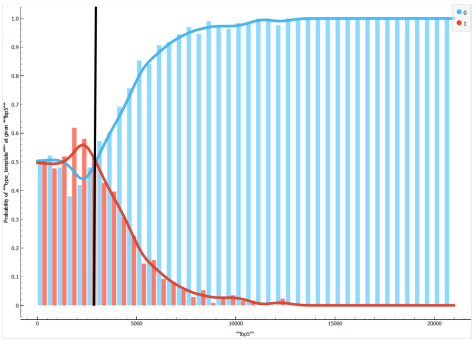

Fig. 8. Conditional probability for a specific feature.

- Multi-features analysis: For the most informative features, we can plot all samples in the training dataset given their values. Figure 9 shows an example for an unknown sample represented in black. In this case, it is quite clear that the sample corresponds to an attack (as it is part of the red area associated to attacks). This is visual indicator easy to understand.

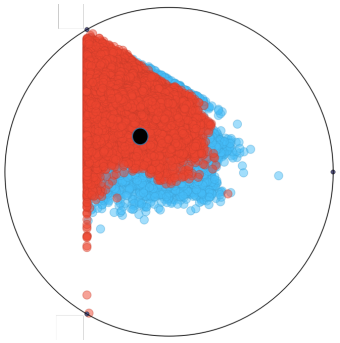

Fig. 9. Projection on the subspace of the most 3 reliable features of an unknown sample represented as a black dot (red points correspond to attack samples, blue to bona fide).

- Models confidence: Any machine learning model can return a probability value for each class. For a PAD system, it is useful to compute a confidence measure by considering this value or when different models are used the strengh of the consensus between us. Figure 10 shows the output of each model on the sample in Figure 7. All agree that this sample is an attack. We can easily compute a confidence value by averaging the probability values (by weighting or not the efficiency of each model). Another possible feedback is the number of models agreeing on the output (in this case, 7/7).
- PAI recognition: In case of attack, the presentation attack instrument (PAI) can be identified and given as feedback to the operator. Figure 11 presents the performance of recognition. The meta model (stack) provides the best result with an accuracy of 98.8% , using a neural network achieves alone an accuracy 98.7%.

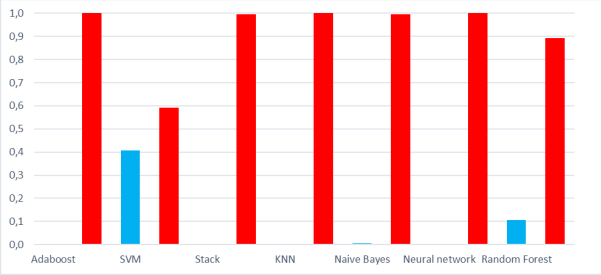

Fig. 10. Example of confidence values for one attack sample. Each model outputs 1 as an attack.

Model	AUC	CA	F1	Prec	Recall	MCC
AdaBoost	0.952	0.952	0.952	0.953	0.952	0.904
SVM	0.584	0.529	0.517	0.548	0.529	0.084
Stack	0.999	0.988	0.988	0.988	0.988	0.976
kNN	0.965	0.959	0.959	0.959	0.959	0.917
Naive Bayes	0.698	0.629	0.627	0.644	0.629	0.276
Neural Network	0.998	0.987	0.987	0.987	0.987	0.973
Random Forest	0.996	0.975	0.974	0.975	0.975	0.949

Fig. 11. PAI recognition results.

VI. Conclusion and perspectives

We showed in this work that we can propose a very efficient PAD system (less efficient than deep PAD systems such as [26] with a CA value of 99.5%) but providing useful and easy to understand feedback to an operator to explain the decision. We intend as perspective to use CNN as feature extraction to enhance results and generate other explanations (attention map as for example).

References

[1] J. Galbally, J. Fierrez, R. Cappelli, and G. L. Marcialis, "Introduction to presentation attack detection in fingerprint biometrics," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pp. 3–15, Springer, 2023.

[2] Joannes Falade, Sandra Cremer, and Christophe Rosenberger, "Comparative study of fingerprint database indexing methods," in *2019 Cyberworlds Conference*, Dec. 2019.

[3] R. Tolosana, M. Gomez-Barrero, C. Busch, and J. Ortega-Garcia, "Biometric presentation attack detection: Beyond the visible spectrum," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1261–1275, 2019.

[4] K. Shaheed, P. Szczuko, M. Kumar, I. Qureshi, Q. Abbas, and I. Ullah, "Deep learning techniques for biometric security: A systematic review of presentation attack detection systems," *Engineering Applications of Artificial Intelligence*, vol. 129, p. 107569, 2024.

[5] H. Sun, Y. Zhang, P. Chen, H. Wang, and R. Liang, "Internal structure attention network for fingerprint presentation attack detection from optical coherence tomography," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2023.

[6] R. Casula, M. Micheletto, G. Orrú, G. L. Marcialis, and F. Roli, "Towards realistic fingerprint presentation attacks: The screenspoof method," *Pattern Recognition Letters*, vol. 171, pp. 192–200, 2023.

[7] E. Marasco and A. Ross, *A Survey on Anti-Spoofing Schemes for Fingerprint Recgnition Systems*. ACM Computing Surveys, Vol. 47,, 2014.

[8] E. Marasco and C. Sansone, *Combining perspiration and morphology based static features for fingerprint liveness detection*. Pattern Recognition Letters, 2012.

[9] K. Raja, R. Ramachandra, S. Venkatesh, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "Vision transformers for fingerprint presentation attack detection," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pp. 17–56, Springer, 2023.

[10] M. Trokielewicz, A. Czajka, and P. Maciejewicz, "Presentation attack detection for cadaver iris," in *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, pp. 1–10, IEEE, 2018.

[11] Z. Qi, S. Khorram, and L. Fuxin, "Embedding deep networks into visual explanations," *Artificial Intelligence*, vol. 292, p. 103435, 2021.

[12] S. Booth, Y. Zhou, A. Shah, and J. Shah, "Bayes-trex: a bayesian sampling approach to model transparency by example," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 11423–11432, 2021.

[13] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pp. 80–89, IEEE, 2018.

[14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[15] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, 2016.

[16] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

[17] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pp. 818–833, Springer, 2014.

[18] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, "Smoothgrad: removing noise by adding noise," *arXiv preprint arXiv:1706.03825*, 2017.

[19] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PloS one*, vol. 10, no. 7, p. e0130140, 2015.

[20] W. J. Von Eschenbach, "Transparency and the black box problem: Why we do not trust ai," *Philosophy & Technology*, vol. 34, no. 4, pp. 1607–1622, 2021.

[21] A. Kortylewski, B. Egger, A. Schneider, T. Gerig, A. Morel-Forster, and T. Vetter, "Empirically analyzing the effect of dataset biases on deep face recognition systems," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 2093–2102, 2018.

[22] P. C. Neto, T. Gonçalves, J. R. Pinto, W. Silva, A. F. Sequeira, A. Ross, and J. S. Cardoso, "Explainable biometrics in the age of deep learning," *arXiv preprint arXiv:2208.09500*, 2022.

[23] K. Abhishek and A. Yogi, *A minutiae Count Based Method for Fake Fingerprint Detection*. ScienceDirect, 2015.

[24] J. Demšar and B. Zupan, "Orange: Data mining fruitful and fun-a historical perspective," *Informatica*, vol. 37, no. 1, 2013.

[25] V. Mura, L. Ghiani, G. L. Marcialis, F. Roli, D. A. Yambay, and S. A. Schuckers, "Livdet 2015 fingerprint liveness detection competition 2015," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–6, Sep. 2015.

[26] A. Popli, S. Tandon, J. J. Engelsma, and A. Namboodiri, "A unified model for fingerprint authentication and presentation attack detection," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pp. 77–99, Springer, 2023.

VII. Acknowledgments

This work was supported by a French government grant managed by the Research National Agency under the France 2030 program, reference ANR-22-PECY-0011.